Hierarchical Segment Support for Categorical Image Labeling

Michael Donoser
Institute for Computer Graphics and Vision
Graz University of Technology, Austria
donoser@icg.tugraz.at

Hayko Riemenschneider
Computer Vision Lab (CVL)
ETH Zurich, Switzerland
hayko@vision.ee.ethz.ch

Abstract

This paper introduces a novel method for categorical image labeling, where each pixel is uniquely assigned to one of a set of unordered, discrete labels. Starting from provided label-depending pixel likelihoods we (a) exploit a segment hierarchy as spatial support to define powerful priors and (b) introduce an efficient and effective inference method, that can be implemented in a few lines of code. Experiments show that competitive labeling accuracy compared to related discrete, continuous, segmentation and filtering approaches is achieved.

1. Introduction

Important challenges in computer vision like interactive image segmentation and pixel-wise class recognition can be formulated as a categorical image labeling problem, i. e. the goal is to uniquely assign each pixel of an image to one of a set of pre-defined, unordered and discrete labels. In general, solutions to such labeling problems are found in a two step process. First, likelihoods for assigning individual pixels to the pre-defined labels are obtained in a dense manner. Since selecting the label with the highest likelihood independently yields noisy and not well aligned results, mostly priors encoding knowledge about reasonable label configurations are analyzed in a second step.

The simplest form of prior information is to assume that neighboring pixels should get the same label, thus such approaches only analyze pairwise pixel label assignments. Such pairwise priors are mostly integrated by considering random field formulations, where image labeling is defined as an energy minimization problem, and the lowest energy labeling is returned as final solution. These energy minimization methods in discrete [1] and continuous variants [8] have proven to be an indispensable tool in computer vision. Nevertheless, searching for the optimal solution in such a high-dimensional solution space usually requires complex optimization algorithms. Furthermore, standard pairwise random fields have limited expressive power, since only interactions between pairs of random variables are modeled. These limitations were for example addressed in [5], where a class of higher order potentials denoted as P^n Potts model was introduced, that can be minimized using standard move making approaches. Recently it was also shown that patch based priors lead to excellent results. In [9], it was demonstrated that image labeling results can be efficiently obtained by smoothing the label costs with a fast edge preserving filter.

In this work, we advocate the usage of segment hierarchies to define a powerful spatial prior for the task of categorical image labeling. The basis of our approach is a hierarchy of segments at different granularity, which is obtainable from any provided gradient magnitude image. We furthermore show, that based on this segmentation hierarchy we are able to introduce a novel, effective inference approach to provide highly accurate labeling results. Our inference process locally selects segments of optimal granularity from the segment hierarchy in an efficient manner and (a) yields smooth labeling results, which (b) obey the local label likelihoods and (c) have aligned label transitions with high contrast locations. In such a way, our approach avoids the problem of selecting segments of suitable granularity as required in higher-order random fields, provides more reasonable spatial support in comparison to patch-based approaches and circumvents the problem of segment boundary over-smoothing as in standard pairwise random fields.

Our method is mainly related to three recent papers: (a) the pylon model proposed in [6], (b) the consistency graph based approach of [4] and (c) the unsupervised segmentation method of [2]. We would like to emphasize the main differences to these approaches.

In [6], authors proposed a labeling method analyzing segment hierarchies for the specific task of semantic segmentation. The core idea of their work is that given a segment tree, semantic segmentation can still be formulated using standard, pairwise random field formulations, which are then solvable using efficient and effective graph cut algorithms like alpha expansion. Our proposed method differs in following points: our basic data structure is a multi-level, multi-granularity hierarchy instead of an unbalanced segmentation tree built by agglomerative clustering. Additionally, instead of applying multi-label graph cuts to infer a solution, we propose to use a parameter-free, graph theoretical approach denoted as Maximum Weight Independent Set (MWIS) for inference, which is implementable in a few lines of code. Therefore, we do not require any additional solver like alpha expansion.

In [4] a large pool of different figure-ground segmentations of the image is used to build a so-called consistency graph. A set of valid segment compositions (tilings) are then found as maximal cliques in the consistency graph. These tilings may induce residual regions, that have to be handled separately. Labeling is independently done for each tiling candidate using belief propagation, and the labeling with the highest probability is finally selected. Our proposed method again differs in several points: first, while [4] generates several, unrelated tilings, we define a segment hierarchy, which avoids required parametrization for obtaining segments at different granularity. Second, in [4], each tiling candidate has to be labeled independently, whereas we propose a joint segmentation and labeling approach. Third, in contrast to [4] our hierarchical segment tree guarantees that all pixels are labeled, avoiding the required cumbersome handling of residual regions. Finally, our inference can be coded in a few lines of code, again avoiding the need of an additional solver like Belief propagation.

In [2], the problem of unsupervised segmentation was addressed using a similar inference process as in our method. Core idea is to select the optimal subset from a large pool of segmentation candidates by applying a Maximum Weight Independent Set (MWIS) algorithm. Since the returned segments may not be able to cover the entire image again some kind of post-processing is required. Our proposed method first differs in the application, since we address image labeling and not unsupervised segmentation, which leads to different formulations for the unary and pairwise terms. Second, in contrast to [2], our MWIS step is based on a hierarchical data structure that guarantees complete covering of the image, and in such a way we do not require any cumbersome post-processing.

2. Method

Our method solves a Maximum a Posteriori assignment problem analyzing cliques of random variables, similar to the wide-spread random field formulations. Our cliques are defined within a hierarchical data structure representing segmentation results at different granularity, as it is described in Section 2.1. Section 2.2 then defines our Maximum a Posteriori labeling problem and introduces our novel inference method based on solving a Maximum Weight Independent Set (MWIS) problem.

2.1. Segment Hierarchy

The first step of our proposed image labeling method is to build a hierarchical segment data structure based on the edge features that are used to define the contrast sensitive Potts models in pairwise random fields. We assume that the edge features are provided in terms of a gradient magnitude image ∇I. We accomplish this step by building a graph structure that is based on the well-known data structure denoted as component tree.

Let $\mathrm{X} \in \mathbb{R}$ be the domain on which the gradient magnitude image $\nabla I : \mathrm{X} \rightarrow [0,1]$ is defined. The domain range is equally split into D increasing values that define the evolution space $t_1, t_2, \ldots t_D$ with $t_d \in [0,1]$ and $t_d < t_{d+1}$, where t_d is a threshold variable. For every value t_d we can define a so-called cross-section F_{t_d} of ∇I by

$$F_{t_d} = \{ i \in \mathcal{V} \mid \nabla I\left(i\right) \leq t_d \} \ , \tag{1}$$

i. e. each cross section contains the pixels having a gradient magnitude value below the corresponding threshold value. We then identify all connected components $C_s^{t_d}$ of the cross-sections F_{t_d}, which represent segments in the image enclosed by gradient magnitude values that are higher or equal as the corresponding t_d value. These connected components (segments) $C_s^{t_d}$ constitute the basis of our segment hierarchy, where each segment becomes a single node of the tree. By increasing the value t_d, the segments can only become larger, and therefore a segment at level t_d is always entirely included in a single segment at level t_{d+1}, which defines unique inclusion relationships between the nodes. Observing the evolution of the individual segments from the leaf nodes to the root node of the tree, we see components appear (in the leaves) and merge (within the tree), until we obtain a single segment consisting of the entire image in the root node. In such a way we define a hierarchical segment tree representation containing segmentations of different granularity, down from the root node (entire image) to the leaf nodes, which represent some kind of superpixels. Furthermore, each level of the tree corresponds to one of the cross-sections F_{t_d}.

The obtained tree is now extended with additional edges to define a unique graph structure $\mathcal{G}$ which is used for the subsequent inference step, that is described in detail in the next section. Our graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a set of nodes $\mathcal{V}$ and a set of edges $\mathcal{E}$. The nodes are the obtained segments $C_s \in \mathcal{V}$ (we will skip the threshold value for notational simplicity in the rest of the paper). An edge $e = (C_s, C_t)$ with $e \in \mathcal{E}$ connects *any* pair of nodes C_s and C_t that overlap each other, i. e. that have a path (not crossing the root node) connecting them within the initial tree structure. Note that in such a way, all overlapping segments are connected, independent of the number of levels between them. We further assume from now on that the number of segments (nodes) within the tree equals S.

2.2. Image Labeling Inference

The segment hierarchy $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ obtained by the steps described in the previous section now constitutes the basis for our efficient image labeling inference step. We additionally assume that we have given pixel-wise values $p\left(x_i\right)$, representing the likelihoods for assigning a pixel to a specific label. How to obtain these likelihoods depends on the application, and details on our choices are given in the experiments. Our novel inference strategy aims at assigning each pixel in the image to a unique node of the hierarchy (a segment) by maximizing the decision certainty considering the provided likelihoods. Our overall goal is to select the maximally distinctive set of non-overlapping segments from the hierarchy, that uniquely partitions the image, and in such a way directly defines the final labeling result. We propose to obtain this labeling solution by solving the so-called Maximum Weight Independent Set (MWIS) problem on our graph $\mathcal{G}$.

Consider our graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with the set $\mathcal{V}$ of nodes (segments) and the set $\mathcal{E}$ of edges obtained in the way as described in the previous section. Assume further, that each node (segment) $C_s \in \mathcal{V}$ has associated a corresponding cost w_{C_s}. A subset of $\mathcal{V}$ can be represented by a binary vector $\mathbf{y} = \{y_1, y_2, \ldots y_S\}$, where S is the overall number of segments in the hierarchy and $y_i = 1$ if y_i is in the subset, and $y_i = 0$ otherwise. A subset $\mathbf{y}$ is denoted as *independent set* if no two nodes of $\mathbf{y}$ are connected by an edge, i. e. $\forall\left(C_s, C_t\right) \in \mathcal{E} \ \mid \ y_s + y_t \neq 2$. A *Maximum Weight Independent Set* (MWIS) is the independent set with the lowest sum over its corresponding node costs.

Similar to standard random fields, this problem can be formulated in a probabilistic manner, which aims at finding the Maximum a Posteriori solution by maximizing

$$p\left(\mathbf{y}\right) = \frac{1}{Z} \prod_{C_s \in \mathcal{V}} \exp\left(-w_{C_s}\, y_s\right) \prod_{\left(C_s, C_t\right) \in \mathcal{E}} \mathbf{1}_{\left(y_s + y_t \neq 2\right)} \quad , \tag{2}$$

where $\mathbf{1}$ is an indicator function and $\mathbf{1}_{(1)} = 1$ and $\mathbf{1}_{(0)} = 0$. As can be seen if $\mathbf{y}$ is not an independent set, $p\left(\mathbf{y}\right)$ becomes 0, which guarantees that the finally activated segments ($y_i = 1$) are non-overlapping. Thus our sought Maximum Weight Independent Set $\mathbf{y}^*$ can be found by

$$\mathbf{y}^* = \underset{\mathbf{y}}{\operatorname{argmax}}\, p\left(\mathbf{y}\right) \quad . \tag{3}$$

We define the costs w_{C_s} for each segment by considering the underlying certainty of the label assignments within each segment. In a first step we build a pixel-wise uncertainty map $H\left(i\right)$ measuring the entropy of the pixel-wise labeling probabilities $p\left(x_i\right)$ by

$$H\left(i\right) = -\sum_{x_i \in \mathcal{L}} p\left(x_i\right)\ log\ p\left(x_i\right) \ , \tag{4}$$

where i is again the index for the pixels in the image. The costs for the segments C_s are then defined by calculating an entropy weighted average of the pixel log likelihoods within the segment by

$$w_{C_s} = \sum_{i \in C_s} -\, log\, p\left(x_i\right)\ H\left(i\right) \ . \tag{5}$$

In such a way, we get low costs for segments consisting of pixels that all have a high assignment likelihood to the same label, i. e. where the labeling decision is quite certain. We are now able to find an approximate solution of the MAP assignment problem by applying an MWIS algorithm on the defined weighted graph structure $\mathcal{G} = (\mathcal{V}, \mathcal{E}, w_{C_s})$. The MWIS algorithm uniquely assigns each pixel to one of a set of non-overlapping segments, i. e. it finds the MAP solution of Equation 2, and by assigning the corresponding MAP label to each segment, we obtain the final labeling result.

Solving the MWIS problem is known to be NP hard, but numerous approximation algorithms exist. We apply one of the simplest, very easy to implement methods: the evolutionary replicator dynamics [11]. Replicator Dynamics can be implemented in a few lines of code in every programming language and variants for efficient optimization in linear time are available [3]. Additionally, these dynamics guarantee to converge to a local maximum.

3. Experiments

We evaluate our method for interactive segmentation and make a thorough comparison to related continuous, discrete, filtering- and segmentation based labeling methods. We use the multi-label interactive segmentation benchmark denoted as *IcgBench* [10], which provides seeds and ground truth segmentation results for 262 images. Segmentation quality is evaluated by the so-called dice score [10] which measures the amount of mutual overlap between the analyzed segmentation and the ground truth.

Since we aim at directly demonstrating the contribution of the proposed categorical image labeling method, we used the same label likelihoods (provided by [10]) as input to all labeling methods compared. These likelihoods were calculated by a Random Forest classifier evaluated on several color and texture features (see [10] for more details) and the pairwise potentials were defined be the gradient magnitude estimator of [7]. Table 1 compares the obtained dice scores to the maximum likelihood result (no regularization) and results obtained by related methods [10, 1, 9, 6]

As can be seen in Table 1 our method achieves the highest dice score of $\mathbf{92.78}\%$. The improved performance (+2.03%) compared to the local filtering method proposed in [9] especially demonstrates the importance of considering the powerful spatial support of image segments compared to the fixed quadratic window support used in [9].

IcgBench	MaxLike	SuperPixel	GC [1]	TV-L1 [10]	Filtering [9]	Pylon [6]	**Proposed**
Dice	84.72	87.60	92.14	92.49	90.75	91.44	**92.78**

Table 1: Dice scores for different image labeling methods on the interactive segmentation benchmark *IcgBench* [10].

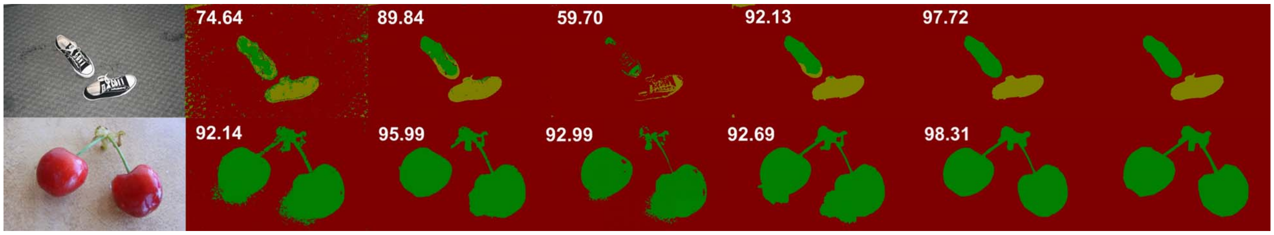

Figure 1: Selected results and corresponding labeling accuracy for interactive segmentation [10]. First column shows input image, second maximum likelihood assignment, third Graph Cut regularization [1], fourth Cost Filtering [9], fifth Pylon regularization [6], sixth our proposed method and seventh ground truth. As can be seen our approach e. g. circumvents the problem of segment boundary over-smoothing as in standard pairwise random fields (see cherry) and clearly demonstrates the usefulness of considering spatial support defined by a segment hierarchy.

If we only consider the superpixel representation (the leaf nodes of the segment hierarchy) for regularization and assign the maximum likelihood label to each superpixel segment, a dice score of only 87.60% is achieved. This furthermore clearly demonstrates the effectiveness of selecting segments on individual levels of the hierarchy. Although the obtained improvements might seem marginal, even small differences in pixel-wise accuracy can produce a significant difference in the segmentation quality as also can be seen in Figure 1, which shows qualitative results, comparing the analyzed labeling methods on selected images.

4. Conclusion

In this paper we introduced a novel, efficient and effective categorical image labeling method. We described how to build a segment hierarchy to define a powerful spatial prior for labeling. We then defined a label certainty score for each segment in the hierarchy, which was afterwards analyzed in a Maximum Weight Independent Set (MWIS) algorithm. The MWIS yields a unique, smooth labeling result which obeys the given labeling likelihoods and has label transitions at high image contrast locations. Experimental evaluation demonstrated that we achieve promising performance for an interactive segmentation task, especially circumventing well-known problems of standard pairwise random fields like segment boundary over-smoothing.

References

[1] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 23, 2001.

[2] W. Brendel and S. Todorovic. Segmentation as maximum-weight independent set. In *Proc. of Advances in Neural Information Processing (NIPS)*, pages 307–315. 2010.

[3] S. R. Bulò and I. M. Bomze. Infection and immunization: a new class of evolutionary game dynamics. *Games and Economic Behaviour*, 71:193–211, 2011.

[4] A. Ion, J. Carreira, and C. Sminchisescu. Probabilistic joint image segmentation and labeling. In *Proc. of Advances in Neural Information Processing (NIPS)*. 2011.

[5] P. Kohli, M. P. Kumar, and P. H. S. Torr. P3 & beyond: Solving energies with higher order cliques. In *Proc. of Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[6] V. S. Lempitsky, A. Vedaldi, and A. Zisserman. Pylon model for semantic segmentation. In *Proc. of Advances in Neural Information Processing (NIPS)*, pages 1485–1493, 2011.

[7] M. Maire, P. Arbelaez, C. Fowlkes, and J. Malik. Using contours to detect and localize junctions in natural images. In *Proc. of Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[8] T. Pock, A. Chambolle, D. Cremers, and H. Bischof. A convex relaxation approach for computing minimal partitions. In *Proc. of Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[9] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz. Fast cost-volume filtering for visual correspondence and beyond. In *Proc. of Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[10] J. Santner, T. Pock, and H. Bischof. Interactive multi-label segmentation. In *Proc. of Asian Conf. on Computer Vision (ACCV)*, 2010.

[11] J. W. Weibull. *Evolutionary game theory*. Cambridge University Press, 1995.