

Image Segmentation by Cascaded Region Agglomeration

Zhang, Liqiang

June 24, 2018

Abstract

In this paper, the author propose a hierarchical segmentation algorithm that starts with a very fine oversegmentation and gradually merges regions using a cascade of boundary classifiers. This approach allows the weights of region and boundary features to adapt to the segmentation scale at which they are applied. The stages of the cascade are trained sequentially, with asymmetric loss to maximize boundary recall. On six segmentation data sets, our algorithm achieves best performance under most region-quality measures, and does it with fewer segments than the prior work.

1. Introduction

A standard preprocessing step in many recognition tasks today is to partition the input image into a set of superpixels: “perceptually meaningful atomic regions”. In a typical vision system these number in the hundreds. Sometimes a coarser partition is used, with only tens (or perhaps just a handful) of regions; in this regime the regions are no longer atomic, but the hope is that they remain perceptually meaningful, that is, that each region does not straddle boundaries between semantically distinct regions, such as boundaries of an object, or occluding boundaries. Thus, such image partition is often called oversegmentation.

Regions extracted by oversegmentation usually form a representation of an image that is much more compact than the original pixel grid. As long as the oversegmentation indeed does not undersegment any region, this speeds up reasoning, with little loss of accuracy for the downstream task, such as category-level segmentation or depth estimation. A user of oversegmentation algorithm has two conflicting objectives: On the one hand, make the superpixels as large as possible, hence reducing their number; on the other hand, preserve true boundaries in the image, hence driving the number of superpixels up and their size down. All the state-of-the-art superpixel methods provide a tuning parameter that controls this tradeoff, and the range that seems to work for a variety of applications is 400 to 1000

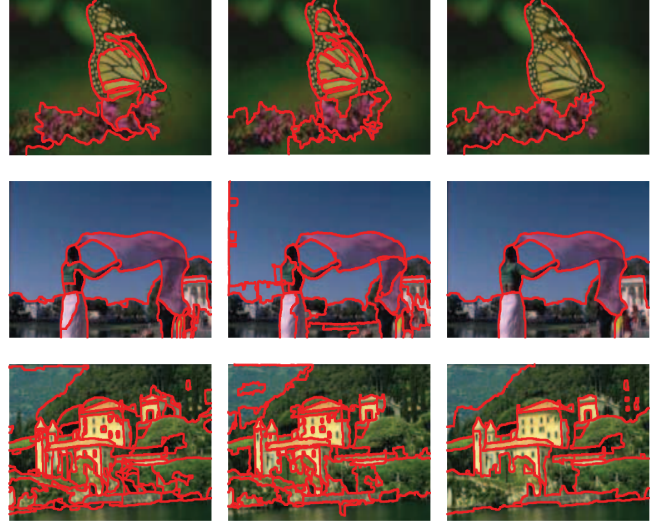


Figure 1. Example segmentations by OWT-UCM (left), Hoiem *et al.* [2] (middle) and ISCR (right). For each image, segmentation shown is at the scale optimal for the image (OIS) with each algorithm, the best that algorithm can do for the image.

superpixels.

This team show that across six data sets, performance measures and for a broad range segmentation scale from fine to coarse, the performance of ISCR is superior to that of current state-of-the-art methods. It is by far the most comprehensive evaluation of the leading approaches to superpixel extraction, and in general to image segmentation. Some typical examples for ISCR compared to leading other methods are shown in Fig. 1.

2. Problem setup

The author model Pg with linear logistic regression as Eq. 1:

$$Pg(p, q; I, \mathcal{R}) = 1/(1 + \exp\{-w^T \phi_{pq}(I, \mathcal{R})\}). \quad (1)$$

The model is usually trained by minimization of log-loss, averaged over examples (which here means averaged over region pairs in the training set). However, naive use of

logloss fails here, because of the symmetry with which it penalizes two types of mistakes: over- and underestimating Pg . Given a typical image with 1000 superpixels in \mathcal{R} , only a small fraction of pairs in $N(R)$ will be true negatives, since most neighboring superpixels should be grouped together. A similar problem was identified in [4], leading to a modification of Hamming loss. They scale the loss value for every pair of regions by the length L_i , in pixels, of the boundary between the regions. This scaling reflects the higher loss from “erasing” a long true boundary than a short one:

$$w * (a) = \underset{w}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n L_i [y_i \log Pg(\phi_i; w) + a(1 - y_i) \log(1 - Pg(\phi_i; w))]. \quad (2)$$

The scaled loss in Eq. 2 is convex and optimizable in the same way as the “normal” logistic regression.

3. Results on benchmarks

Method	Covering		VOI		PRO	
	OIS	ODS	OIS	ODS	OIS	ODS
UCM [1]	0.65	0.59 ²³	1.54	1.66 ¹⁰	0.85	0.81 ⁵⁰
Hoiem [3]	0.59	0.55 ¹⁷	1.70	1.83 ¹⁰	0.82	0.79 ¹⁷
ISCRA	0.66	0.60 ¹²	1.40	1.61 ⁸	0.86	0.81 ¹⁶

Table 1. Results for large segment regime, **BSDS300**.

Method	Covering		VOI		PRO	
	OIS	ODS	OIS	ODS	OIS	ODS
UCM [1]	0.64	0.59 ¹⁸	1.49	1.69 ¹³	0.85	0.83 ³⁹
Hoiem [3]	0.60	0.56 ¹⁸	1.66	1.78 ¹¹	0.84	0.8 ¹⁸
ISCRA	0.66	0.59 ²⁴	1.42	1.60 ¹⁰	0.85	0.82 ²⁴

Table 2. Large segment regime, **BSDS500**.

Method	Covering		VOI		PRO	
	OIS	ODS	OIS	ODS	OIS	ODS
UCM [1]	0.74	0.64 ⁶	1.05	1.30 ³	0.85	0.78 ¹²
Hoiem [3]	0.67	0.65 ⁷	1.34	1.37 ⁷	0.80	0.77 ⁸
ISCRA	0.75	0.67 ⁴	1.02	1.18 ³	0.85	0.77 ¹⁴

Table 3. Large segment regime, **MSRC**.

Results for region measures relevant for the large segment regime are summarized in Tables 1 through 3.

References

- [1] P. Arbellez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE TPAMI*, 33(5):898–916, 2011. 2
- [2] D. Hoiem, A. A. Efros, and M. Hebert. Geometric context from a single image. In *ICCV*, 2005. 1
- [3] D. Hoiem, A. A. Efros, and M. Hebert. Recovering occlusion boundaries from an image. *IJCV*, 91(3):328–346, 2011. 2
- [4] S. Kim, S. Nowozin, P. Kohli, and D. Y. Chang. Higher-order correlation clustering for image segmentation. In *NIPS*, 2011. 2