Segmenting natural images with the least effort as humans

Qiyang Zhao
http://gait.buaa.edu.cn/~zqy

State key lab of software development environment,
Beihang University

Given that natural image segmentation is well-known as an ill-posed problem, then how can we design an algorithm to obtain good performance as human subjects? A choice is to learn from human segmentations as MCG [3] which obtains the state-of-the-art performance and fairly high computational efficiency. Then a question arises here: what should we learn? The way all existing literatures exploit human segmentations ignores a basic fact that human segmentations are produced by human operations. The human segmentation process would inevitably fulfill the human behavior's principles including **the least effort principle (LEP)**: a human *will strive to solve his problem in such a way as to minimize the total work that he must expend* [1].

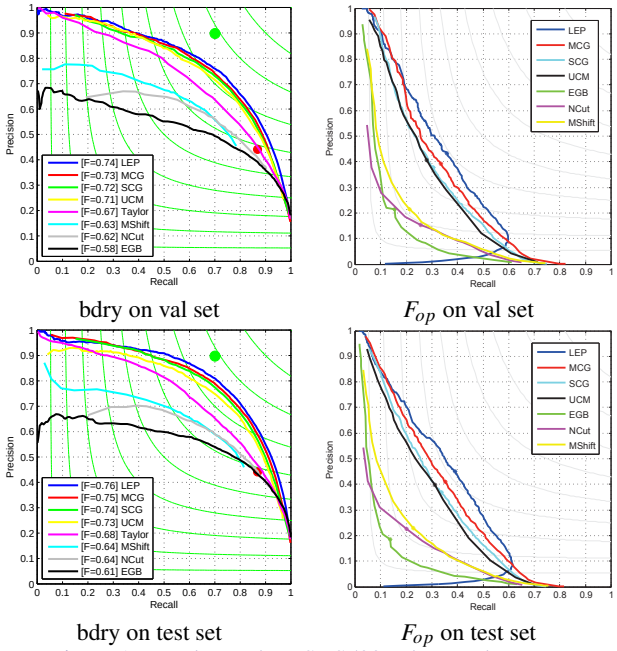
The principle gives us a new insight into our problem. Suppose you are a human subject in the BSDS segmentation experiment, and are required to segment images into pieces under the **ending instruction**: *each piece contains only one single distinguished thing*. Then your total effort F of segmenting an image should including understanding images (U) to guide following operations, and tracing boundaries (T) by hands with a mouse: $F(I,S) = U(I) + T(S)$. Then from the viewpoint of LEP, human subjects would like to find an acceptable segmentation S by minimizing their effort, or formally,

$$\min_S F(I,S), \text{s.t. each piece contains only one thing.} \quad (1)$$

We can describe a region with features/concepts with colors, textures and semantic concepts. If you can describe a region easily, then there tend to be only one 'thing'. But the description effort is not the only factor we should take into account when deciding whether to partition a region or not, instead we must balance the description effort and the tracing effort. In order to estimate the tracing effort T, we estimate the tracing loads in an empirical way. We collect several human subjects to trace the boundaries in the BSDS human segmentations, and tell them to do it as precisely as possible. The tracing style is to use many small clicks as recommended by the BSDS tools. We record the time consumptions as the measure of tracing loads, then analyze boundaries to find a way to predict them. Finally we find the tracing loads of long boundaries are linearly related to the corner amounts of different angle bins, while it is different for short boundaries as the loads seem to be only linearly related to lengths. We mix these two cases softly with a Logistic function to get a unified model. The term U is hard to tackle, because analyzing it with either the psychology or the neural science is far beyond the scope of the paper. Instead, we simply set it as a constant value.

The BSDS human segmentations are produced hierarchically with the BSDS tool. In order to mimic human operations, we sort it into an **Hierarchy Constraint** on our model, and prove that the optimal solutions on larger unit costs can be found by merging the region group of the minimum *muc* in the optimal solutions on smaller unit costs. However, the amount of all region groups is too huge to be handled. A feasible way is to consider adjacent region pairs only, so to optimize F approximately with a naive merging strategy. The key idea is to maintain a dynamically sorted search tree *BST*, but all adjacencies should be updated and re-inserted in *BST* after each mergence. By profiling we find about 80% updates are wasted as they are overwritten by new updates. Fortunately, we find almost all mergences are subject to a **Monotonicity Constraint**, with which we prove that the output of the naive merging strategy is the exact solution minimizing F, and the *muc*'s are monotonically increased. In light of the monotonicity, we replace the search tree *BST* in naive merging with a direct access table to improve the computational efficiency considerably.

There are several popular measures to evaluate the segmentation algorithms independently [2]. However in existing literatures, the best performances of a single method on these measures are usually reported on dramatically different thresholds. We integrate them to get a more objective measure which can reflect both boundary and region performance synchronously. We train our new segmentation method on the BSDS train set and evaluate our new segmentation method (LEP for abbr.) on the test set and val set. We calculate the color entropy in the Lab space and the texture entropy based on a variant of C-LBP, to measure the independent descriptions, and use normalized Euclidean Lab- distances and the odds being separated in the structured predictions of a random forest to measure the regularity description. We evaluate all involved methods with their public source codes or pre-computed results. The experiment on B-SDS500 shows that our method obtains the state-of-the-art performance, as shown in figures and tables. Especially on the test set, our method is the only one with the precision 0.90+ on the recall 0.50 (see fig.1.a). It runs on all images in less than 1s in average, and is much more efficient than other method which has 0.70+ scores on the boundary measure.



bdry on val set F_{op} on val set

bdry on test set F_{op} on test set

Figure 1: Results on the BSDS500 val set and test set.

Table 1: Results on the BSDS500 val set. The best values are in bold.

	Boundary			F_{op}		Covering		PRI		VOI		Time
	ODS	OIS	AP	ODS	OIS	ODS	OIS	ODS	OIS	ODS	OIS	
Human	0.79	0.79	-	-	-	0.73	0.73	0.87	0.87	1.16	1.16	-
Ours	**0.74**	**0.76**	**0.79**	**0.39**	**0.46**	**0.62**	**0.68**	**0.82**	**0.86**	**1.51**	**1.31**	1s
MCG	0.73	0.76	0.74	0.36	0.42	0.61	0.67	0.81	0.86	1.55	1.37	20s+
SCG	0.72	0.73	0.63	0.34	0.40	0.59	0.65	0.80	0.85	1.62	1.44	3.5s+

Table 2: Results on the BSDS500 test set. The best values are in bold.

	Boundary			F_{op}		Covering		PRI		VOI		Time
	ODS	OIS	AP	ODS	OIS	ODS	OIS	ODS	OIS	ODS	OIS	
Human	0.80	0.80	-	0.56	0.56	0.72	0.72	0.88	0.88	1.17	1.17	-
Ours	**0.76**	**0.79**	**0.82**	**0.42**	**0.47**	**0.63**	**0.69**	**0.84**	**0.87**	**1.47**	**1.29**	1s
MCG	0.75	0.78	0.76	0.38	0.43	0.61	0.66	0.83	0.86	1.57	1.39	20s+
SCG	0.74	0.77	0.65	0.35	0.42	0.60	0.65	0.83	0.86	1.63	1.43	3.5s+

The success of our new algorithm shows, when learning human segmentations, it is feasible and necessary to take human behaviors into account to find a good solution. We also observe that the performance of our algorithm on the new integrating measure is much worse than that of human subjects. How to introduce semantic concepts into our algorithm is our future work. All source codes and evaluation results are available on http://gait.buaa.edu.cn/~zqy/lep.

[1] Zipf G. *Human behaviour and the principle of least effort*. Addison-Wesley Press, 1949.

[2] Pont-Tuset J. and Marques F. Measures and meta-measures for the supervised evaluation of image segmentation. In *CVPR*, 2013.

[3] Arbelaez P., Pont-Tuset J., and Barron J. et al. Multiscale combinatorial grouping. In *CVPR*, 2014.