Tri-Map Self-Validation Based on Least Gibbs Energy for Foreground Segmentation

Xiaomeng Wu
wu.xiaomeng@lab.ntt.co.jp
Kunio Kashino
kashino.kunio@lab.ntt.co.jp

NTT Communication Science Laboratories
3-1, Morinosato Wakamiya Atsugi-shi
Kanagawa, Japan 243-0198

Foreground segmentation plays an important role in high-level vision tasks. Of previously reported research, a large percentage is made up of Markov random field (MRF) based studies [2, 5, 6], in which optimal segmentation maximizes the posteriori probability given observations incorporated with a predefined tri-map. They are current to the state-of-the-art, but under the assumption that a sufficiently discriminative tri-map is given, *e.g.* specified by user interaction [6] or supervised by using class information [2, 5]. With a low-quality tri-map, although some attempts have been made to improve the MRF model, very little attention has been paid to enhancing the discernment of the tri-map itself. This constitutes the main problem that we tackle in this paper.

In contrast to the previous studies, which depended on strong assumptions, our aim is *unsupervised* foreground segmentation under only one weak (realistic) assumption. We assume that the location of a foreground is a normal deviate in the image space, whose expectation lies near the center of the image. We argue that the least Gibbs energy (LGE) can be formulated as a goal function of a tri-map optimization problem, and propose decomposing the complex problem into a series of tractable subproblems. A suboptimal optimization is gradually obtained by making decisions between pixel cluster-level set operations.

(a) $LGE = 2.61 \times 10^6$ (b) $LGE = 2.47 \times 10^6$ (c) $LGE = 2.42 \times 10^6$

Figure 1: Different tri-maps (left) exhibit differences in least Gibbs energies (LGE), incorporated in the segmentation (right) of the same image.

In terms of MRFs, the optimal segmentation $\hat{X}$ maximizes the a posteriori probability pertaining to an observed image Y and a tri-map T. It is equivalent to minimizing the Gibbs energy $E(X|Y,T)$:

$$E(X|Y,T) = \sum_p \sum_\alpha U_p^{(\alpha)}(y_p|T)\delta(\alpha, x_p) + \sum_{p,q} \frac{1-\delta(x_p,x_q)}{\|p-q\|}\exp(-\beta\|y_p - y_q\|) \quad (1)$$

where the right terms are known as the likelihood (first) and coherence (second) energies at the pixel level. We define the LGE as follows:

$$LGE(T|Y) = \min_X E(X|Y,T) \quad (2)$$

LGE is a function of T with a given observation Y, and is no longer dependent on the segmentation X. When the distributions of foreground and background pixels offer very low separability, as shown in Fig. 1(a), the likelihood term becomes non-contributory and the minimization over-fits the coherence term, resulting in a high LGE. When tri-maps lead to the same segmentation, i.e. to equivalent coherence energies, as shown in Fig. 1(b) and 1(c), the tri-map with the larger distribution overlap indicates a higher entropy. A desired tri-map $\hat{T}$ can be defined as one that minimizes $LGE(T|Y)$, more specifically

$$\hat{T} = \arg\min_T \min_X E(X|Y,T) \quad (3)$$

We propose a split-and-validate method for solving this problem. The splitting is determined by a non-parametric clustering method (see the paper). After splitting, the image is abstracted as a set of pixel clusters. Our tri-map validation is based on two types of cluster-level operations:

(Retaining) *Keeping a tri-map T unchanged, as denoted by $T \leftarrow T$.*

(Contracting) *For a tri-map $T = \{T_B, T_F\}$, in which T_B and T_F are background and foreground regions, and a pixel cluster c, subtracting c from T_F and adding c to T_B, as denoted by $T \leftarrow \{T_B \cup c, T_F \setminus c\}$.*

The self-validation of a tri-map is discretized to a tree-structured evolution process. $T^{(0)}$ is preliminarily treated as a rectangle in the center. Using Eq. 1, we can obtain $LGE(T^{(0)}|Y)$. All pixel clusters $\{c_1, c_2, \cdots\}$ are sorted in ascending order of image-space centrality. This is motivated by the assumption that a cluster of pixels is more likely to belong to the foreground if its location is closer to the center of the image. $T^{(0)}$ is then arguably refined by **Contracting** with the cluster at the top of the sorted queue, which leads to a tentative tri-map $T'^{(0)}$ and $LGE(T'^{(0)}|Y)$. An arbitrary T is contract-able if **Contracting** leads to a lower LGE than **Retaining**. If so, we update T to T' and continue this process iteratively until all clusters are incorporated in the validation. We obtain the segmentation by using an iterated graph cut [6] with the refined $\hat{T}$.

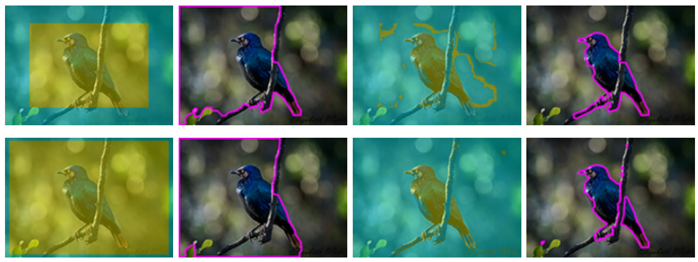

Figure 2: Example of tri-map optimization and segmentation. From left to right: initialized tri-map, segmentation of GC [6], optimized tri-map, and our segmentation.

Figure 2 compares the segmentations initialized by the same tri-map. Table 1 compares our method with advanced studies. More detail regarding the non-parametric clustering method determining the splitting and the experiments is described in the paper. Our conclusion is that the LGE can be a strong cue for capturing the discriminative power of a tri-map, and is useful when dealing with unsupervised foreground segmentation.

Table 1: Performance on Oxford Flower17 reported in the literature[1].

Method	MJI	MNHS
Nilsback and Zisserman [5]	93.0	–
Joulin *et al.* [3]	75.8	86.6
Chai *et al.* [2]	**94.7**	**98.3**
Najjar and Zagrouba [4]	84.0	–
Aydin and Ugur [1]	87.0	–
Suta *et al.* [7]	90.0	89.0
Our Method	**91.7**	**96.8**

[1] The definition of *MJI* and *MNHS* can be found in the paper.

[1] D. Aydin and A. Ugur. Extraction of flower regions in color images using ant colony optimization. *Procedia CS*, 3:530–536, 2011.

[2] Y. Chai, V. S. Lempitsky, and A. Zisserman. BiCoS: A Bi-level co-segmentation method for image classification. In *ICCV*, pages 2579–2586, 2011.

[3] A. Joulin, F. R. Bach, and J. Ponce. Discriminative clustering for image co-segmentation. In *CVPR*, pages 1943–1950, 2010.

[4] A. Najjar and E. Zagrouba. Flower image segmentation based on color analysis and a supervised evaluation. In *ICCIT*, pages 397–401, 2012.

[5] M.-E. Nilsback and A. Zisserman. Delving deeper into the whorl of flower segmentation. *Image Vision Comput.*, 28(6):1049–1062, 2010.

[6] C. Rother, V. Kolmogorov, and A. Blake. "GrabCut": interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, 2004.

[7] L. Suta, F. Bessy, C. Veja, and M.-F. Vaida. Active contours: Application to plant recognition. In *ICCP*, pages 181–187, 2012.