

Flooding and MRF-based Algorithms for Interactive Segmentation

Ilias Grinias, Nikos Komodakis and Georgios Tziritas

Computer Science Department, University of Crete, Heraklion, Greece
{grinias, komod, tziritas}@csd.uoc.gr

Abstract

We propose a method for interactive colour image segmentation. The goal is to detect an object from the background, when some markers on object(s) and the background are given. As features only probability distributions of the data are used. At first, all the labelled seeds are independently propagated for obtaining homogeneous connected components for each of them. Then the image is divided in blocks, which are classified according to their probabilistic distance from the classified regions. A topographic surface for each class is obtained, using Bayesian dissimilarities and a min-max criterion. Two algorithms are proposed: a regularized classification based on the topographic surface and incorporating an MRF model, and a priority multi-label flooding algorithm. Segmentation results on the LHI data set are presented.

I. Introduction

Natural image segmentation is one of the most challenging problems in image analysis. Often the objective is to extract the foreground object out of the cluttered background. The automatic segmentation problem is ill-posed and complex, as more than one partition can satisfy generic objective criteria. Unambiguous solutions, or segmentations satisfying subjective criteria, could be obtained, if the user draws labelled scribbles on the regions of interest and on the background.

In an interactive segmentation framework the user provides prior information as initial scribbles, and sometimes intervenes to correct intermediate inaccurate partitions. In [1] hard topological constraints are used to reduce the search space of feasible segmentations minimizing an energy function via max-flow/min-cut algorithms. In [2] weighted geodesic distances to

user-provided scribbles are computed, from which the image is automatically segmented.

In this paper, we suppose that the user provides a few points defining scribbles of approximately circular shape with two possible labels: foreground and background. In the first step connected components of homogeneous image features are obtained, by the independent local extension of the scribbles, via a minimum spanning tree. The image data are represented by the distribution of colour components. From these homogeneous connected components are obtained the features for as many classes as the number of initial scribbles. The unlabelled parts of the image are then classified in a block-wise way to one of the initial classes. Having the block-wise classification, we obtain the distributions for all labelled classes. Then, probabilistic distances are used to determine and classify pixels that belong to one of the classes with high confidence, including the hard constrained scribbles. Having available the data modelling and the initial map of classified pixels, we use flooding algorithms in order to compute a topographic surface per class by assigning a height to each unclassified pixel, using Bayesian dissimilarity criteria. Finally, the local pixel interactions of label heights are modelled by a *Markov random field (MRF)* model and the underlying energy is minimized by a *primal-dual* algorithm [3]. The algorithms employed in this paper for interactive segmentation, have been adequately adapted from those introduced in [4] for automatic colour and texture segmentation.

II. Model Learning

For each initial scribble of circular shape a lattice graph is constructed with only horizontal and vertical links. Having an initial vertex for all initial scribbles, the minimum spanning tree (MST) is constructed on the graph using Prim's algorithm. The link weight

is defined by Mallows distance separately on three data components: Y for the intensity and (a, b) for the chrominance. As the data sets have equal size the Mallows distance is calculated after sorting the data for each component and using a *city-block* distance [5]. From the distances on the MST a threshold is determined, assuming that at most 30% of the links in the tree could correspond to boundary crossings. In addition a second threshold is determined based on testing the homogeneity of the grouping around the initial scribble. If one of the two thresholds is exceeded the tree is cut and the component containing the initial seed is retained.

At the end of the first stage a part of the image is classified. From the classified data are obtained the probability density functions, always separately for the three colour components. A feature for measuring local variation is added, computed by a Laplacian filter. The unclassified part is divided in blocks, which are classified to one of the classes according to a similarity criterion, based again on the Mallows distance. However, in the case of blocks not sufficiently similar to any label and lying near the image border, new classes may be introduced on the background. Finally, the whole image data are classified and the description of the classes is obtained using probability distributions.

Therefore a probabilistic model for the image data has been learned. Let s denote a pixel site, and $\xi(s)$ denote the vector of features at this site. As the result of the previous process we have obtained $p_l(\xi(s))$, the conditional probability density function of class l . Finally, taking into account the probability distributions for all classes, we obtain the *a posteriori* probabilities

$$\Pr\{l|\xi(s)\} = \frac{P_l p_l(\xi(s))}{\sum_{k=1}^L P_k p_k(\xi(s))}, \quad (1)$$

where P_k is the *a priori* probability of class k among L possible. The *a priori* probabilities can be estimated in some configurations of initial scribbles, otherwise they are assumed all equal.

III. Flooding Process for Class Propagation

A. Class Initialization

We seek sets of spatially connected regions of pixels, which are classified to class l with high confidence, using statistical tests. For each pixel s a classification is performed based on the maximum *a posteriori* probability in Eq. (1). Then the connected components for each class are determined and the larger components are retained and get classified.

These initial regions of high confidence may be used for the construction of a minimum spanning tree for each class of the image. Indeed, metric

$$d_l^B(s) = -\ln \Pr\{l|\xi(s)\}$$

could be used for defining the edge weight of the tree attributed to class l . At the same time the initial regions could be interpreted as topographic valleys of minimum height in a topographic map, where the height represents dissimilarity.

B. A min-max classification criterion

Let $S = \bigcup_{l=1}^L S_l$ be the set of those initially classified pixels. For any unclassified pixel s we can consider all the paths linking it to a classified set or region. A path $C_l(s)$ is a sequence of adjacent pixels $\{s_0, \dots, s_n\}$, where $s_n = s$, while all pixels of the sequence are unlabelled, except s_0 which has label l . The cost of a particular path is defined as being equal to the maximum cost of a pixel classification according to the Bayesian rule and along the path

$$\max_{i=1, \dots, n} d_l^B(s_i).$$

Therefore, for each l a topographic surface on a discrete grid is defined, considering 4-connected pixels. The initially classified pixels are defined to be at the zero level, while the height of the unclassified pixels is given by the Bayesian rule.

Finally, the classification problem becomes equivalent to search for the shortest path under the above cost, as we can define the distance of any unclassified pixel from the different classes as being the lowest height to climb for reaching site s ,

$$\delta_l(s) = \min_{C_l(s)} \max_{s_i \in C_l(s)} d_l^B(s_i). \quad (2)$$

Therefore the decisions are topologically constrained.

If we consider the graph of unclassified sites with 4-connections and the classified connected components, we can define an edge weight as follows

$$w(s_{i-1}, s_i) = \max(d_l^B(s_{i-1}), d_l^B(s_i)).$$

Paths defined by Eq. (2) belong to the MST of the graph defined above and the computation of $\delta_l(s)$ necessitates the construction of this MST [4].

On the other hand, it is very interesting to remark that the classification problem, as posed here, consists of constructing a topographic surface, as that for finding watershed lines [6]. Hence, we can use a region growing procedure, like the immersion (flooding) algorithm [7] for computing the above defined heights

and distances and for classifying pixels, taking into account region features and topology constraints.

In Subsections III-C and III-D two algorithms, based on the principle of the *min-max* Bayesian criterion for classification, are presented in detail. The main difference between them is that in the former topological constraints are limited to the connectivities imposed by the *min-max* criterion and the final classification is left to a MRF-based minimization method, while in the second strong topological constraints are imposed during the classification process.

C. MRF-Based Minimization on the Topographic Map (ILFMA)

At first distances $\delta_l(s)$ of Eq. (2) are computed by L independent flooding procedures. The initially classified pixels, composing spatially connected regions, are grown by iteratively considering neighboring pixels. The growing procedure is applied to compute the distances $\delta_l(s)$. Among all neighboring pixels to the set S_l , that are unclassified and of unknown distance from label l , the nearest pixel is found, according to Eq. (2). Growing proceeds until no more pixels can be added to the expanding regions, because their propagating contour reaches only pixels with different initial labels.

Given the region growing measurements derived in Eq. (2), we then propose to optimize a discrete *MRF* in order to decide what the final class labels should be. In this manner, we aim at capturing the local interactions between pixels, which will help us to refine and correct the class labels that were assigned during the previous stage of the algorithm. The problem can be formulated as follows: we seek to assign a class label l_s (from a discrete set of class labels \mathcal{L}) to each node of a lattice graph $s \in \mathcal{V}$, so that the following cost is minimized:

$$\sum_{s \in \mathcal{V}} c(l_s) + \sum_{(s,z) \in \mathcal{E}} w(l_s, l_z), \quad (3)$$

where \mathcal{E} is the set of the graph edges. The singleton potentials are set according to the dissimilarity measurements derived in Eq. (2), *i.e.*, $c(l_s) = \delta_l(s)$, while the pairwise potentials are set according to the Potts function, with all weights w set equal to a constant, adapted to the dissimilarities to be regularized, in case of different classification and zero potential when the two neighboring points are assigned the same class.

For minimizing the MRF energy in (3), we make use of the recently proposed *primal-dual* method in [3], which casts the MRF optimization problem as an integer program and then makes use of the duality theory of linear programming in order to derive solutions that have been proved to be almost optimal.

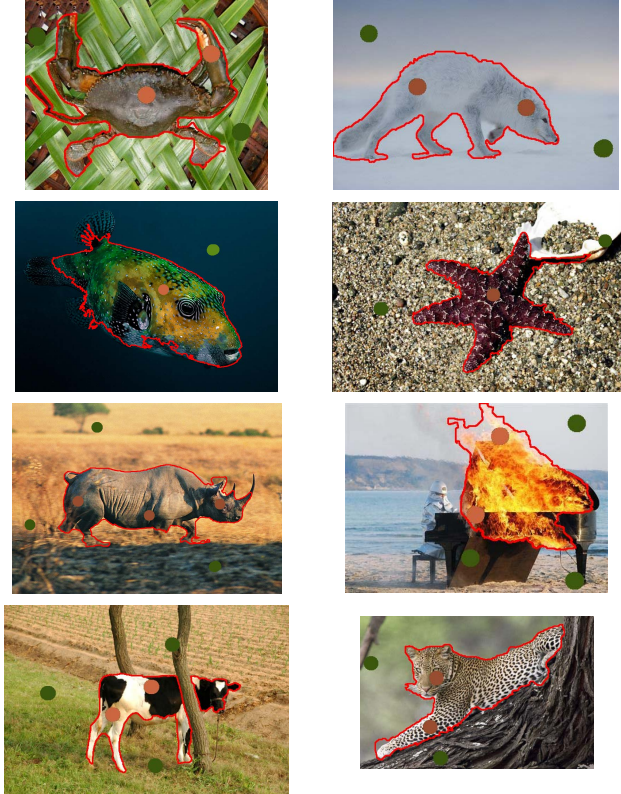


Figure 1. Segmentation results for eight images of the *LHI Interactive Segmentation Dataset*. From left to right and from top to bottom: 0,1, 4, 8, 9, 20, 22, 35. The user marked scribbles are also shown.

D. Priority Multi-Class Flooding Algorithm

The **Priority Multi-Class Flooding Algorithm (PMCFA)** imposes strong topology constraints. All initially classified regions are propagated simultaneously and most likely decisions are taken as soon as possible. All the contours of initially classified regions are propagated towards the space of unclassified image pixels, according to similarity criteria, which are based on the class label and the segmentation features. Contour pixels s are sorted according to their dissimilarity $\delta_{l(R)}(s)$ from the class label $l(R)$ of regions R they adjoin and at each step, a *group* of contour pixels of minimum dissimilarity are set to the label of the class to which they most probably belong.

In addition, group classification refers to pixels s , which are placed on region contours at a step of the propagation progress and whose metric $\delta_l(s)$ against class label l is similar. This fact, implies the *quantization* of distance metric $d_l^B(\cdot)$, which in turn

No	RI	Vol	BDE	RI	Vol	BDE
0	0.93	0.44	3.2	0.93	0.46	3.1
1	0.95	0.34	3.3	0.94	0.40	4.1
2	0.98	0.13	1.2	0.98	0.13	1.2
4	0.87	0.60	15.6	0.87	0.60	16.2
6	0.95	0.31	2.9	0.95	0.31	2.8
8	0.95	0.30	8.6	0.95	0.30	8.9
9	0.97	0.23	5.2	0.97	0.22	4.6
10	0.98	0.12	1.3	0.99	0.12	1.3
14	0.90	0.54	10.1	0.89	0.56	10.3
15	0.97	0.22	2.7	0.97	0.24	2.7
18	0.93	0.45	4.6	0.93	0.47	4.8
20	0.95	0.33	3.8	0.95	0.32	3.6
21	0.95	0.28	12.0	0.96	0.23	9.8
22	0.97	0.18	2.0	0.97	0.19	2.0
23	0.90	0.48	12.4	0.80	0.91	20.9
24	0.98	0.16	3.7	0.98	0.16	4.4
25	0.87	0.72	8.0	0.86	0.77	8.0
26	0.98	0.16	2.4	0.97	0.18	3.3
34	0.87	0.64	18.7	0.87	0.66	32.0
35	0.96	0.29	2.3	0.95	0.35	2.7
36	0.96	0.28	3.6	0.96	0.28	3.6
Mean	0.94	0.34	6.1	0.93	0.37	7.2

Table I. Performance evaluation

leads to the reduction of spatial redundancy that often appears during growing. Spatial coherence arises when neighboring pixels have almost equal distance value or equivalently, as it is clear by the definition of $d_l^B(\cdot)$, when neighboring pixels have almost equal posterior probabilities for a class given their feature vector.

Furthermore, two decision thresholds, namely \mathcal{T}_l^{high} and \mathcal{T}_l^{low} , may be defined on the metric. A pixel s surely belongs to class l , if $\Pr\{l|\xi(s)\} > 0.5$. We therefore set $\mathcal{T}_l^{high} = \ln 2$. A pixel s does not belong to class l , if $\Pr\{l|\xi(s)\} < \frac{1}{L}$. Thus, we could set $\mathcal{T}_l^{low} \geq \ln L$. The implementation of the flooding algorithm uses quantization of distances and priority lists to obtain a quasi-linear computational complexity.

IV. Experimental Results and Conclusions

We have tested our method on the LHI interactive segmentation benchmark [8]. We tested only the scribbles considered as the hard level of difficulty. Results on eight images are given in Fig. 1 obtained by ILFMA. In addition, three metrics are employed to evaluate the precision on foreground/background labelling and boundary localization: Rand index (1 is the best), Variation of Information (0 is the best) and Boundary Displacement (0 is the best). The definitions of the metrics can be found in [9]. The results for the whole benchmark are given in Table 1, in the left-side columns for the ILFMA and in the right-side columns for the PMCFA. Both algorithms give good to very good results for this benchmark. PMCFA is

faster, while ILFMA provides more regular regions and smoother boundaries.

In conclusion, an interactive image segmentation framework has been proposed. Only probability distributions of the image data are used. The model learning is based on propagating the given scribbles and obtaining homogeneous and topologically coherent clusters. The clusters define the image data classes. The model learning is the more time-consuming module, but determinant for the segmentation performance. Then, a topographic surface is computed per class, using Bayesian dissimilarity criteria. For the final image segmentation two algorithms are proposed: an *MRF*-based minimization and a priority multi-class flooding algorithm. Possible improvements include the stage of model learning and the incorporation of boundary constraints using the image gradient.

References

- [1] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient N-d image segmentation," *Intern. Journal of Computer Vision*, vol. 70, pp. 109–131, 2006.
- [2] X. Bai and G. Sapiro, "Geodesic matting: A framework for fast interactive image and video segmentation and matting," *Intern. Journal of Computer Vision*, vol. 82, pp. 113–132, 2009.
- [3] N. Komodakis and G. Tziritas, "Approximate labeling via graph cuts based on linear programming," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 1436–1453, Aug. 2007.
- [4] I. Grinias, N. Komodakis, and G. Tziritas, "Bayesian region growing and MRF-based minimization for texture and colour segmentation," in *Intern. Workshop on Image Analysis for Multimedia Interactive Services*, 2007.
- [5] E. Levina and P. Bickel, "The earth mover's distance is the Mallows distance: Some insights from statistics," in *Proc. of the IEEE Intern. Conf. on Computer Vision*, vol. 2, 2001, pp. 251–256.
- [6] F. Meyer, "Topographic distance and watershed lines," *Signal Processing*, vol. 38, pp. 113–125, 1994.
- [7] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 583–598, June 1991.
- [8] The LHI data set. [Online]. Available: <http://www.imageparsing.com/interactivesegmentation.html>
- [9] M. Krinidis and I. Pitas, "Color texture segmentation based on the modal energy of deformable surfaces," *IEEE Trans. on Image Processing*, vol. 18, pp. 1613–1622, 2009.