Interactive Image Segmentation via Graph Clustering and Synthetic Coordinates Modeling

Costas Panagiotakis[1], Harris Papadakis[2], Elias Grinias[3], Nikos Komodakis[4], Paraskevi Fragopoulou[2], and Georgios Tziritas[5]

[1] Dept. of Commerce and Marketing, Technological Educational Institute (TEI) of Crete, 72200 Ierapetra, Greece,
cpanag@staff.teicrete.gr,
[2] Dept. of Applied Informatics and Multimedia, TEI of Crete, PO Box 140, Greece,
adanar@epp.teicrete.gr, fragopou@ics.forth.gr**
[3] Dept. of Geoinformatics and Surveying, TEI of Serres, 62124 Serres, Greece,
elgrinias@gmail.com
[4] Dept. of Computer Science and Applied Math, Ecole des Ponts ParisTech, France,
nikos.komodakis@enpc.fr
[5] Dept. of Computer Science, University of Crete, P.O. Box 2208, Greece,
tziritas@csd.uoc.gr

Abstract. We propose a method for interactive image segmentation. We construct a weighted graph that represents the superpixels and the connections between them. An efficient algorithm for graph clustering based on synthetic coordinates is used yielding an initial map of classified pixels. The proposed method minimizes a min-max Bayesian criterion that has been successfully used on image segmentation problem taking into account visual information as well as the given markers. Experimental results and comparisons with other methods demonstrate the high performance of the proposed scheme.

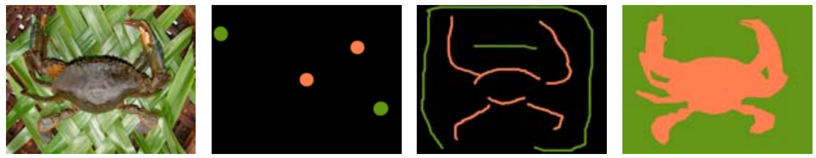
Keywords: image segmentation; network coordinates; graph clustering

1 Introduction

Image segmentation is a key step in many image-video analysis and multimedia applications. According to interactive image segmentation, which is a special case of image segmentation, unambiguous solutions, or segmentations satisfying subjective criteria, could be obtained, since the user gives some markers on the regions of interest and on the background. Fig. 1 illustrates an example of an original image, two types of markers and the segmentation ground truth.

During the last decade, a large number of interactive image segmentation algorithms have been proposed in the literature. In [1], a new shape constraint based method for interactive image segmentation has been proposed using Geodesic paths. The authors introduce Geodesic Forests, which exploit the structure of

** Paraskevi Fragopoulou is also with the Foundation for Research and Technology-Hellas, Institute of Computer Science, 70013 Heraklion, Crete, Greece.



(a) (b) (c) (d)

Fig. 1: **(a)** Original image, **(b), (c)** given markers and **(d)** the ground truth image.

shortest paths in implementing extended constraints. In [2], discriminative learning methods have been used to train conditional models for both region and boundary based on interactive scribbles. In the region model, the authors use two types of local histograms with different window sizes to characterize local image statistics around a specific pixel. In the boundary model, the authors use 12-bin boundary features by applying gradient filters to each color component. In [3], a two step segmentation algorithm has been proposed that first obtains a binary segmentation and then applies matting on the border regions to obtain a smooth alpha channel. The proposed segmentation algorithm is based on the minimization of the Geodesic Active Contour energy.

According to the interactive segmentation algorithm proposed in [4], first, all the labeled seeds are independently propagated for obtaining homogeneous connected components for each of them. Then, the image is divided into blocks which are classified according to their probabilistic distance from the classified regions, and a topographic surface for each class is obtained. Finally, two algorithms for regularized classification based on the topographic surface have been proposed.

The proposed method can be divided into several steps. In the first step, we partition the image into superpixels using the oversegmentation algorithm proposed in [5]. Then, we construct a weighted graph that represents the superpixels and the connections between them, taking into account the given markers and visual information (see Sections 2). Next, we use the Vivaldi algorithm [6] that generates the superpixels' synthetic coordinates (see Section 3). An initial map of classified pixels is provided by an efficient algorithm for graph clustering based on synthetic coordinates (see Section 4). Thus, we solve the graph clustering problem using the synthetic network coordinates that are automatically estimated by a distributed algorithm based on interactions between neighboring nodes. Finally, the image segmentation is provided by a Markov Random Field (MRF) model or a flooding algorithm minimizing a min-max Bayesian criterion (see Section 5). Hereafter we present the proposed methodology, a more detailed analysis of the proposed method is given in [7].

2 Graph Generation

Initially, we partition the image into superpixels using the oversegmentation algorithm proposed in [5]. In this work, the description of visual content consists of Lab color components for color distribution and textureness for texture content. This approach has been also used in [8]. The visual distance $d_v(s_i, s_j)$ between two superpixels s_i and s_j is given by the Mallows distance [9] of the three color components in Lab color space and for the textureness measure of the corresponding superpixels. Let G' be the weighted graph of superpixels, so that two superpixels s_i and s_j are connected with an edge of weight $d_v(s_i, s_j)$ if and only if they are neighbors, meaning that they share a common boundary. Then, the proximity distance $d_p(s_i, s_j)$ between superpixels s_i and s_j is given by the length of the shortest path from s_i to s_j in graph G'. The proposed distance between superpixels s_i and s_j that efficiently combines the visual and proximity distances is given by Equation 1:

$$d(s_i, s_j) = \sqrt{d_p(s_i, s_j)} \cdot d_v(s_i, s_j) \tag{1}$$

The use of the square root on the proximity distance is explained by the fact that the visual distance is more important than the proximity distance. The graph G' is used in order to compute the graph G that is defined hereafter.

In the next step, we construct a graph G that represents the superpixels and the connections between them, taking into account the given markers and visual information. According to the given markers, two superpixels can either be connected, meaning that they belong to the same class or be disconnected, meaning that they belong to different classes. Thus, the nodes (superpixels) in this graph are connected with edges of two types:

- the E_C edges that connect two superpixels belonging to the same class,
- the E_D edges that connect two superpixels belonging to different classes,

taking into account the two types of relations between superpixels. In the second step of the algorithm, the visual distance and the superpixels' proximity are used to create the set of edges E_C until G becomes a connected graph.

Hereafter, we present the procedure that computes the two sets of edges, E_C and E_D for graph G. E_C and E_D are initialized to the corresponding edges according to the given markers. Then, the $\frac{N \cdot (N-1)}{2}$ pairs of distances $d(.,.)$ are sorted and stored in vector v, where N denotes the number of superpixels. We add the sorted edges of v on E_C set until G becomes a connected graph in order to be able to execute the Vivaldi algorithm [6] that generates the superpixels' synthetic coordinates (see Section 3). In addition, we keep the graph balanced (almost equal degree per node) using an upper limit on node degree ($MaxConn = 10$).

3 Synthetic coordinates

In this work, we have used Vivaldi [6] to position the superpixels in a virtual space (the $n-$dimensional Euclidean space $\Re^n$, e.g. $n = 20$). Vivaldi [6] is a fully decentralized, light-weight, adaptive network coordinate algorithm that predicts Internet latencies with low error. Recently, we have successfully applied Vivaldi on the problem of locating communities on real and synthetic dataset graphs [10, 11]. In the current work, the input to the Vivaldi algorithm is a weighted graph, where the weights correspond to the nodes distances in $\Re^n$. We have used the weights 0.0 and 1000.0 for E_C and E_D, respectively. These weights correspond to the Euclidean distance between the virtual position of superpixels, that is used by the Vivaldi algorithm to position the superpixels in $\Re^n$ generating synthetic coordinates so that the Euclidean distance of any two superpixel positions approximates the actual distance (edge weight) between those superpixels. This means that superpixels of the same class will be placed closer in space than superpixels of different classes, forming natural clusters in space.

4 Graph Clustering

Having estimated a synthetic coordinate (position) $p(s_i) \in \Re^n$ for each superpixel $s_i, i \in \{1, ..., N\}$ of the graph, we can use a clustering algorithm in order to cluster a subset of superpixels into foreground and background classes, providing this way an initial map of classified pixels Map. The proposed algorithm creates the initial map by merging the superpixels that have been placed in proximity in $\Re^n$, meaning that they should belong to the same class. In this research, we have used a hierarchical clustering algorithm that recursively finds clusters in an agglomerative (bottom-up) mode. We successively merge a cluster c_1 with "UNKNOWN" label with the closest labeled cluster c_2, if the distance between c_1 and c_2 is lower than a predefined threshold T according to their synthetic coordinates. If no such pair of clusters exists, the algorithm terminates. T is automatically computed by the histogram of distances between all points' pairs. T is given by the value that better discriminates the two distributions (distances between points of the same class and between points of different classes) from the histogram of distances.

Usually, it holds that the image borders and especially the pixels close to the four image corners belong to the background class, so we have used the following simple rule so that a pixel is classified to background class if it belongs to an unclassified superpixel and its distance

- from the closest image border is less than 1% of the image diagonal
- from the closest image corner is less than 7% of the image diagonal.

Using the criterion of "unclassified superpixel" the proposed heuristic works well in cases where an object intersects the image boundary. Finally, we perform erosion on the classified superpixels using a disk of 2 pixels radius, in order to be able to correct some boundary errors of the oversegmentation algorithm.

5 Image Segmentation

Hereafter, we briefly describe the image segmentation method. The proposed criterion has been proposed in [4,8]. Let $S = \bigcup_{l=1}^{L} S_l$ be the set of those initially classified pixels estimated by the graph clustering algorithm. 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 $\{s_0, ..., s_{n-1}, s_n = s\}$. It holds that all pixels of the sequence are unlabeled, except s_0 which has label l. The cost of a particular path is defined as the maximum cost of a pixel classification according to the Bayesian rule and along the path

$$Cost(C_l(s)) = \max_{i=1...n} d_l^B(s_i) \tag{2}$$

Finally, the classification problem becomes equivalent to a search for the shortest path given the above cost. Two algorithms based on the principle of the $min-max$ Bayesian criterion for classification, have been used. These algorithms have been proposed in [4, 8].

- According to the Independent Label Flooding MRF-based minimization Algorithm (ILFMA), we use the primal-dual method proposed in [12], 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.
- The Priority Multi-Class Flooding Algorithm (PMCFA), that is analytically described in [8], imposes strong topology constraints. 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.

In what follows, the proposed methods using the MRF model and the flooding algorithm are denoted as SGC-ILFMA and SGC-PMCFA, respectively.

6 Experimental Results

SGC-ILFMA and SGC-PMCFA have been compared with algorithms from the literature according to the reported results of [2] and [13], using the following two datasets:

- The LHI interactive segmentation benchmark [14]. This benchmark consists of 21 natural images with ground-truths and three types of users' scribbles for each image.
- The Zhao interactive segmentation benchmark [13]. This benchmark consists of 50 natural images with ground-truths and four types of users' scribbles (levels) for each image. The higher the level, the more markers are added.

In order to measure the algorithms' performance, we use the Region precision criterion (RP) [2]. RP measures an overlap rate between a result foreground and the corresponding ground truth foreground. A higher RP indicates a better segmentation result.

Fig. 2 depicts intermediate (initial map) and final segmentation results of the proposed methods (SGC-ILFMA and SGC-PMCFA) on an image of LHI dataset (see Fig. 2(a)) and three images of Zhao dataset. We graphically depict the given markers on the original images using red color for foreground and green color for background, respectively. The red, blue and white coloring of intermediate results correspond to foreground, background and unclassified pixels, respectively. Under any case, the results of the proposed algorithms are almost the same, yielding high performance results. In Figs. 2(f), 2(f) and 2(j) the initial marker information suffices for the segmentation. Although in Fig. 2(n) the low number of given markers does not suffice to discriminate the foreground and background classes, the proposed methods give good performance results. A demonstration of the proposed method with experimental results is given in [6].

Using the LHI dataset, we have compared the proposed methods with three other algorithms from the literature: CO3 [2], Unger et al. [3] based on the reported results of [2]. The proposed methods SGC-ILFMA and SGC-PMCFA yield RP 85.4% and 85.2%, respectively, outperforming the other methods. The third and the fourth highest performance results are given by the CO3 [2] method with $RP = 79\%$ and Unger et al. with $RP = 73\%$.

In addition, we have compared the proposed methods with three other algorithms from the literature Couprie et al. [15], Grady [16] and Noma et al. [17] using the reported results of [13] on Zhao dataset. Table 1 depicts the mean region precision (RP) on four different simulation levels of SGC-ILFMA, SGC-PMCFA, Bai et al., Couprie et al., Grady and Noma et al. algorithms. The highest performance results are clearly obtained by the SGC-ILFMA and SGC-PMCFA algorithms, while the third highest performance results are given by the Couprie et al. [15] method that gives similar results with the Grady and Noma et al. method.

Method	**SGC-ILFMA**	**SGC-PMCFA**	**Couprie et al.**	**Grady**	**Noma et al.**
1	66.7%	**68.4%**	50%	46%	49%
2	84.1%	**84.5%**	72%	71%	69%
3	85.3%	**85.7%**	84%	84%	82%
4	88.1%	**88.6%**	88%	88%	87%

Table 1: The region precision (RP) over the Zhao dataset.

7 Conclusion

In this paper, a two-step algorithm is proposed for interactive image segmentation taking into account image visual information, proximity distances as well as the given markers. In the first step, we constructed a weighted graph of superpixels and we clustered this graph based on a synthetic coordinates algorithm. In the second step, we have used a MRF or a flooding algorithm for getting the final image segmentation. The proposed method yields high performance results under different types of images and shapes of the initial markers.

[6] `http://www.csd.uoc.gr/~cpanag/DEMOS/intImageSegmentation.htm`

Acknowledgments

This research has been partially co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Programs: ARCHIMEDE III-TEI-Crete-P2PCOORD, THALIS-NTUA-UrbanMonitor and THALIS-UOA- ERASITECHNIS.

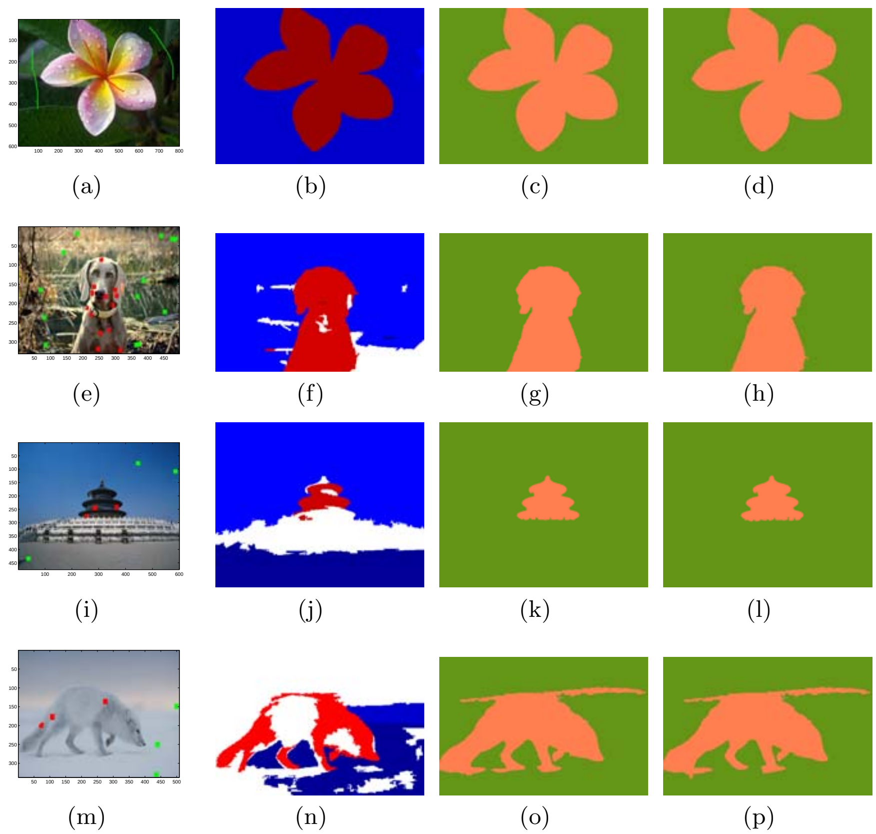


Fig. 2: **(a), (e), (i), (m)** Original images with markers from the LHI and Zhao datasets. **(b), (f), (j), (n)** Initial map of classified pixels. **(c), (g), (k), (o)** Final segmentation results of the SGC-ILFMA method. **(d), (h), (l), (p)** Final segmentation results of the SGC-PMCFA method.

References

1. Gulshan, V., Rother, C., Criminisi, A., Blake, A., Zisserman, A.: Geodesic star convexity for interactive image segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2010) 3129–3136
2. Zhao, Y., Zhu, S., Luo, S.: Co3 for ultra-fast and accurate interactive segmentation. In: Proceedings of the international conference on Multimedia, ACM (2010) 93–102
3. Unger, M., Pock, T., Trobin, W., Cremers, D., Bischof, H.: Tvseg-interactive total variation based image segmentation. In: In: British Machine Vision Conference (BMVC). (2008)
4. Grinias, I., Komodakis, N., Tziritas, G.: Flooding and MRF-based algorithms for interactive segmentation. In: International Conference on Pattern Recognition (ICPR). (2010) 3943–3946
5. Felzenszwalb, P., Huttenlocher, D.: Efficient graph-based image segmentation. International Journal of Computer Vision **59** (2004) 167–181
6. Dabek, F., Cox, R., Kaashoek, F., Morris, R.: Vivaldi: A decentralized network coordinate system. In: Proceedings of the ACM SIGCOMM '04 Conference. Volume 34. (2004) 15–26
7. Panagiotakis, C., Papadakis, H., Grinias, I., Komodakis, N., Fragopoulou, P., Tziritas, G.: Interactive image segmentation based on synthetic graph coordinates. Pattern Recognition (2013 (accepted))
8. Panagiotakis, C., Grinias, I., Tziritas, G.: Natural image segmentation based on tree equipartition, bayesian flooding and region merging. IEEE Transactions on Image Processing **20** (2011) 2276 – 2287
9. Mallows, C.: A note on asymptotic joint normality. The Annals of Mathematical Statistics **43** (1972) 508–515
10. Papadakis, H., Panagiotakis, C., Fragopoulou, P.: Local community finding using synthetic coordinates. In: International Conference on Future Information Technology (FutureTech 2011). (2011)
11. Papadakis, H., Panagiotakis, C., Fragopoulou, P.: Locating communities on real dataset graphs using synthetic coordinates. Parallel Processing Letters (PPL) **20** (2012)
12. Komodakis, N., Tziritas, G.: Approximate labeling via graph cuts based on linear programming. Pattern Analysis and Machine Intelligence, IEEE Transactions on **29** (2007) 1436–1453
13. Zhao, Y., Nie, X., Duan, Y., Huang, Y., Luo, S.: A benchmark for interactive image segmentation algorithms. In: IEEE Workshop on Person-Oriented Vision (POV). (2011) 33–38
14. Yao, B., Yang, X., Zhu, S.: Introduction to a large-scale general purpose ground truth database: methodology, annotation tool and benchmarks. Lecture Notes in Computer Science (**4679**)
15. Couprie, C., Grady, L., Najman, L., Talbot, H.: Power watersheds: A new image segmentation framework extending graph cuts, random walker and optimal spanning forest. In: International Conference on Computer Vision. (2009) 731–738
16. Grady, L.: Random walks for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **28** (2006) 1768–1783
17. Noma, A., Graciano, A., Consularo, L., Cesar-Jr, R., Bloch, I.: A new algorithm for interactive structural image segmentation. arXiv preprint arXiv:0805.1854 (2008)