COMBINATION OF REGION AND CONTOUR MODELS FOR INTERACTIVE IMAGE SEGMENTATION *

Ling Ge, Ran Ju and Gangshan Wu
State Key Laboratory for Novel Software Technology
Nanjing University, Nanjing, China
gelingnju@gmail.com, juran@smail.nju.edu.cn, gswu@nju.edu.cn

ABSTRACT

Previous image segmentation methods are mainly based on region or contours. The former category of models are computationally expensive, while the latter approaches require lots of user interactions. In this paper, we propose a novel interactive image segmentation method, which makes a combination of the two models. By adopting the advantage of respective models, our method can produce high quality segmentations with little user interaction and achieve a surprisingly high efficiency. Specifically, we first obtain a coarse segmentation on a reduced scale using the classical graph cut method. Then we refine the boundary region on finer scales using active contours based method iteratively. The experimental results show that our method can produce better segmentation results to state-of-the-art while greatly reducing user interactions and processing time. We believe the proposed method could greatly improve user experiences in real applications.

Categories and Subject Descriptors

I.4.6 [**Image Processing and Computer Vision**]: Segmentation; I.4.9 [**Image Processing and Computer Vision**]: Applications

General Terms

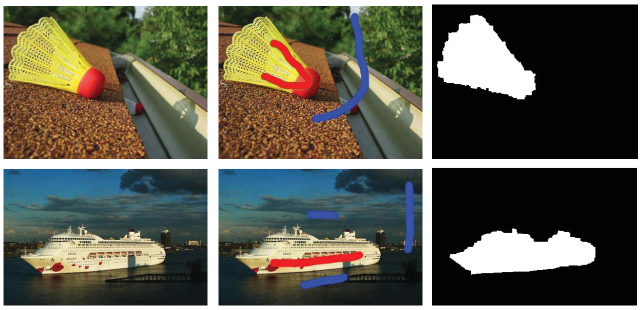
Algorithms

Keywords

Image segmentation, graph cut, active contours, scale space

*This work is supported by the National Science Foundation of China under Grant No.61321491 and No.61202320, Research Project of Excellent State Key Laboratory (No.61223003), Natural Science Foundation of Jiangsu Province (No.BK2012304) and National Special Fund (No.2011ZX05035-004-004HZ).

ICIMCS'15, August 19–21, 2015, Zhangjiajie City, Hunan, China.



(a) (b) (c)

Figure 1: Examples of interactive image segmentation. (a) Original images. (b) Manually labeled seeds of foreground (red) and background (blue). (c) Segmentation results.

1. INTRODUCTION

Interactive image segmentation, which aims to partition images into two parts namely "Object" and "Background" with user interactions as shown in Fig. 1, has attracted much attention in image processing and computer vision society. Typically the problem is formulated as region and contour based models. The former idea solves the segmentation problem as an energy minimization process considering both region statistics and inter-regional relations [5, 11], which usually requires users to mark a few foreground and background regions as seeds to learn the color distribution. The model produces high quality segmentation results but consumes much computation since it performs a second order energy minimization. The second approach works near the object boundaries, which requires users to roughly trace the object contours [10, 3]. Then the local image forces, e.g. gradient, termination, smoothness and so on, and user specified control points are all considered to obtain an optimal solution. Obviously for this approach users have to spend lots of time to supply sufficient interactions .

In this paper, we propose a novel interactive image segmentation method, which combines both the region and contour based models. The proposed method is inspired by the simple but effective observation in daily lives like paper cutting: people always perform the segmentation task by roughly mark the object region, and then spend lots of time in selecting the contours. Similarly, we first perform region based segmentation on a reduced scale of the input image to obtain a coarse result. Then we generate a narrow band near object boundaries according to the initial result, and

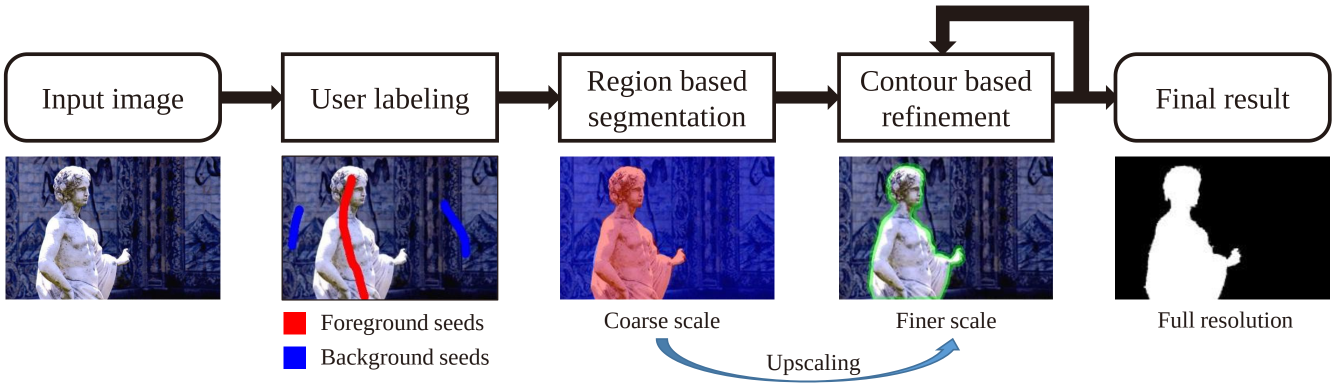


Figure 2: An overview of the proposed method.

perform a contour based refinement on finer scales iteratively. The proposed method has two advantages compared to previous models. First, as the number of pixels for processing is significantly reduced, the computational efficiency is greatly improved. This also enables the proposed method to work for the large photos produced by the high resolution cameras and mobile phones today. Second, compared to the previous contour based methods, our method requires much fewer interactions since the rough contour is generated automatically.

Some previous works [2] use contours or rather edges as constraints for region growing based segmentation. In contrast, we make a sophisticated combination of real object contours and image region attributes in the scale space for interactive image segmentation. There are also some researches dedicated to the efficiency problem of region based methods. For example, in [8] a hierarchical method is proposed to improve the efficiency of graph cut [5]. By representing the image on the scale space, the segmentation is performed in a banded region near object boundaries. The similar idea is also adopted in [9]. Our method is similar to the above in the consideration of multi scales but differs in segmentation models. The previous methods are mainly based on single region model. In contrast, our method makes a combination of region and contour based models, which can produce better segmentation results especially for object contours. Besides, since region models are more expensive than contour models, our method has an improved efficiency than the previous hierarchical models.

The experiments are conducted on the ASD dataset [1] with 1000 images and groundtruth segmentations. We manually mark the foreground and background seeds as inputs. The results compared to state-of-the-art methods show the superior performance in segmentation accuracy and efficiency of the proposed method.

2. MODEL

We give an overview of the proposed method in Fig. 2. Given the input image, the user is asked to mark some regions as foreground and background seeds. Then the region based segmentation is performed on a coarse scale. After that, we employ the active contours based method in the boundary region on finer scales iteratively. Finally, a full resolution result is produced as shown in the rightmost.

2.1 Region Based Coarse Segmentation

For the input image I we first construct its scale pyramid $\{I_0, I_1, I_2, \ldots, I_n\}$, where I_0 represents the full resolution image and I_1 to I_n are successively down sampled from previous level. The number of scale levels is determined as follows. Starting from the original image I_0, we down sample the current scale at a proportion of ρ to construct the next scale. The step is repeated until the pixel number of current scale is below a threshold φ. We can figure out that $n = \lceil log_\rho \frac{|I|}{\varphi} \rceil$.

On the coarsest scale I_n, we employ graph cut [5] to generate the initial segmentation result s_n. The original image I is represented as an undirected graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$, as shown in Fig. 3(a). $\mathcal{V}$ is the set of all pixels with two terminations. The two additional terminal nodes S and T represents foreground and background terminals respectively. $\mathcal{E}$ is the union of neighborhood links (n-links) and terminal links (t-links), where n-links stand for the inter-pixels relation and t-links indicate the individual terminal probability. Then the optimal segmentation can be formulated as a min-cut problem, whose energy function is described as :

$$E_R(L) = \lambda R_R(L) + B_R(L) \tag{1}$$

where $R(L)$ is sum of penalties for assigning a certain pixel node p to foreground and background. $B(L)$ is the sum of penalties for discontinuities between adjacent pixel nodes. L indicates the labels assigned to corresponding pixel nodes, whose value is 1 (foreground) or 0 (background). λ is a parameter to balance the two parts. Specifically,

$$R_R(L) = \Sigma_{p \in I} R_p(l_p) \tag{2}$$

$$B_R(L) = \Sigma_{\{p,q\} \in \mathcal{N}} B_{\langle p,q \rangle} \delta(l_p, l_q) \tag{3}$$

$$\delta(l_p, l_q) = \begin{cases} 0, & if\ l_p = l_q \\ 1, & otherwise \end{cases} \tag{4}$$

where $\mathcal{N}$ is the set of adjacent pixel nodes under a standard 8-neighborbood system. $R_p(L)$ indicates the possibility of pixel p to be labelled as a certain value of L. We compute color distribution of pixels marked by user to get histograms of foreground and background. These histograms are used to calculate the regional penalties $R_p(L)$ as negative log-likelihoods:

$$R_p(L) = -\ln Pr(l_p|L), L \in \{'obj', 'bgd'\}. \tag{5}$$

Boundary penalties $B_{\langle p,q \rangle}$ is defined as

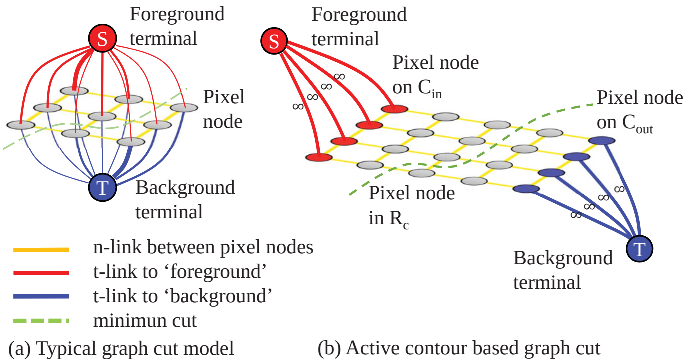


Figure 3: Comparison of different graph cut models.

$$B_{\langle p,q\rangle} = exp(-\frac{(I_p - I_q)^2}{2\sigma^2}), \quad (6)$$

where $(I_p - I_q)^2$ is the Euclidean distance between p and q in RGB color space, and σ is a parameter to adjust the interactions between adjacent pixels.

The minimum energy of Eq. 1 equals to the minimum cut of the graph $\mathcal{V}$, which can be solved using the min-cut algorithm developed by Boykov et al. [4]. After the generation of s_0, we implement an opening operation with a 3×3 element on the result to eliminate noises.

2.2 Active Contours Based Refinement

The initial segmentation result s_0 provides an approximate estimation of foreground and background. Next, we iteratively refine boundary area using a graph cut based active contours method [12]. Suppose that boundary refinement is performed on the kth scale I_k with the initialization of the segment result $s_{(k-1)}$ of the $(k-1)$th scale, the contour based refinement is implemented as follows. First, dilation and erosion operations are performed on $s_{(k-1)}$ to determine the inner contour C_{in} and outer contour C_{out} of foreground respectively. Then $s_{(k-1)}$ is upscaled to the same size of I_k. The region inside C_{in} is set as foreground, the region outside C_{out} is settled as background, and the rest part R_c is treated as the undetermined boundary area. The size of element used in dilation and erosion operations on each scale is set as $(k+2) \times (k+2)$.

Based on the above processing result, we build a new graph $\mathcal{G}_k = \langle \mathcal{V}_k, \mathcal{E}_k \rangle$ as shown in Fig.3(b). The pixels on C_{out} and C_{in} are treated as the source nodes and the sink nodes, respectively. The pixels in R_c are treated as pixel nodes with n-links between them. In this way, we can label the rest pixel nodes using the min-cut algorithm. It is a multi-source multi-sink minimum cut problem. To simplify the problem, we identify the multiple source nodes as a single source node and multiple sink nodes as a single sink node. Note the graph has been converted into a two-dimensional graph, which is different from the 3D case as shown in Fig.3(a). Now the energy function turns to be:

$$E_C(L) = B_C(L) \quad (7)$$

$$B_C(L) = \Sigma_{p \in C_{in} \wedge q \in R_c} B_{\langle p,q\rangle} \delta(1, l_q) + \Sigma_{p \in C_{in} \wedge q \in C_{in}} B_{\langle p,q\rangle} \delta(l_p, l_q) + \Sigma_{p \in R_c \wedge q \in C_{out}} B_{\langle p,q\rangle} \delta(l_p, 0) \quad (8)$$

The energy minimization is only performed on the neighbor link now, which can also be solved using the min-cut algorithm. Since generally there exist sharp gradient changes between object and background and the object contours are naturally smooth, the minimization of Eq. 7 well detects the real object contours. Besides, we may find that the active region for segmentation are greatly reduced, which improves the computational efficiency a lot. Even compared to the previous hierarchical models, our method constructs much less graph edges because terminal links are removed in our contour refinement.

Table 1: Evaluation of different methods.

algorithm	precision	recall	F_β
graphcut	0.8076	0.8161	0.8095
RW	0.8288	0.8448	0.8325
GSC	0.8190	0.8308	0.8217
Ours	0.8327	0.8372	0.8337

3. EXPERIMENTS

3.1 Datasets and Experimental Settings

To evaluate the performance of our method, we select three state-of-the-art methods using the same input types with ours for comparison, namely graphcut [5], Random Walker (RW) [6] and Geodesic Distance (GSC) [7]. The experiments are conducted on the ASD dataset, which includes 1000 images with groundtruth masks. We manually label the foreground and background strokes as default inputs. Throughout the experiment the parameters for our method are set as: $\{\lambda, \varphi, \sigma\} = \{39, 1000, 22\}$. The proposed method is implemented in C++. All the experiments are carried out on a PC with a four-core 3.40GHz CPU and 8GB memory.

3.2 Segmentation Results Evaluation

To evaluate segmentation quality, we use precision, recall and F_β as evaluation metrics, where $\beta^2 = 0.3$ according to [1]. Table 1 shows the comparison results. We may find that our method performs better than the other competitors. Specifically, our method obtains a nearly 3% improvement in F-measure compared to graphcut, which shows that the contour based refinement is reasonable and effective. Besides, we should note that the performances of all the methods are comparable since sufficient supervision has been supplied by users. In Fig. 4 we provide some segmentation results generated by our methods. The original color images, with user marked foreground and background seeds are shown in the first row. The red strokes indicate foreground seeds and blue ones stand for background seeds. Segmentation masks generated by our method are shown in the second row, where white pixels indicate object and black ones are background. We can conclude from the examples that our method could generate accurate segmentation results as well as contour based methods, while being much more convenient in user interaction.

We further give the running time comparison of the compared methods. In this experiment we select 20 images with the resolution over 8M pixels taken by mobile phone indoors and outdoors for evaluation. We resize the images to different resolution and execute algorithms 100 times per image to obtain the average runtime. Fig. 5 shows the average

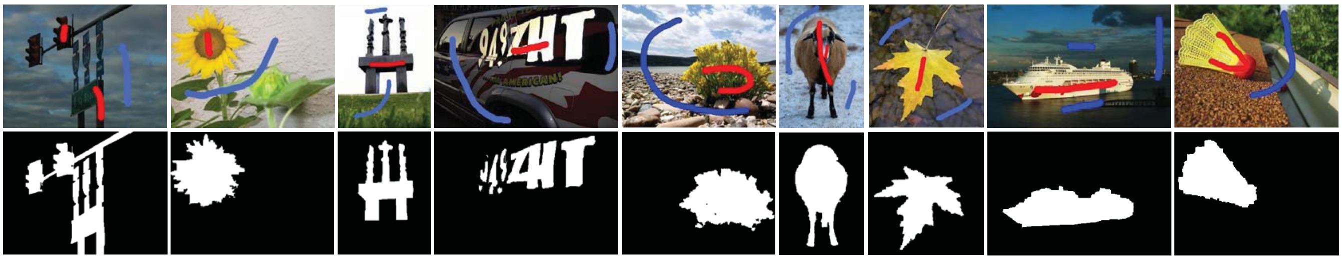


Figure 4: Segmentation results generated by the proposed method. The top row shows the original images with manually labeled strokes, and the second row shows the corresponding segmentation results.

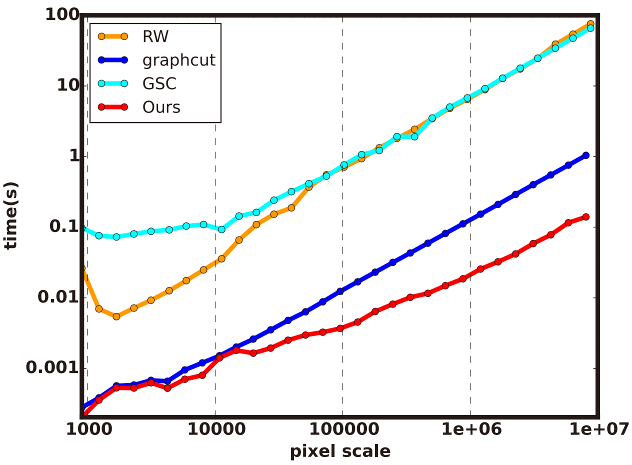


Figure 5: Running time of different methods.

running times of all the methods. It should be noted that we have used LOG-LOG coordinate for illustration since the running times of the benchmark methods stride over three orders of magnitude. We can find that our method shows a significant improvement in efficiency to the other methods owing to the use of coarse-to-fine strategy and contour based refinement. Specifically, our method runs about 7 times faster than graphcut, 296 times faster than GSC, and 341 times faster than RW. A noticeable point in the figure is that the time curves in the leftmost have lower growth rates than in the right, which is mainly caused by the size limitation of cache. The evaluation above also demonstrates that our approach can finish segmenting an 8M resolution image taken by a common mobile phone today in about 0.1 second and thus enables real time user interaction. In contrast, the other methods spend seconds to minutes to perform the task, which is unacceptable for interactive applications.

4. CONCLUSION

In this paper, we proposed a novel method combining both region and contour based models for interactive image segmentation. Owing to the advantages of region and contour based models, our method can generate high quality segmentations with little user interaction, while being much more efficient than previous methods. The high efficiency makes our method work for the high resolution images produced by current portable cameras. Experiments show that the proposed approach can obtain a superior performance compared to state-of-the-art methods in segmentation accuracy and response time, and thus supplies better user experiences for the segmentation task.

5. REFERENCES

[1] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In *CVPR*, pages 1597–1604. IEEE, 2009.

[2] J. Benois and D. Barba. Image segmentation by region-contour cooperation for image coding. In *11th IAPR International Conference on Pattern Recognition, 1992*, pages 331–334. IEEE, 1992.

[3] A. Blake and M. Isard. *Active contours: the application of techniques from graphics, vision, control theory and statistics to visual tracking of shapes in motion*. Springer Science & Business Media, 2012.

[4] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, 2001.

[5] Y. Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in nd images. In *ICCV*, volume 1, pages 105–112. IEEE, 2001.

[6] L. Grady. Random walks for image segmentation. *PAMI*, 28(11):1768–1783, 2006.

[7] V. Gulshan, C. Rother, A. Criminisi, A. Blake, and A. Zisserman. Geodesic star convexity for interactive image segmentation. In *CVPR*, pages 3129–3136. IEEE, 2010.

[8] H. Lombaert, Y. Sun, L. Grady, and C. Xu. A multilevel banded graph cuts method for fast image segmentation. In *ICCV*, volume 1, pages 259–265. IEEE, 2005.

[9] D. Miyazaki, Y. Matsushita, and K. Ikeuchi. Interactive shadow removal from a single image using hierarchical graph cut. In *ACCV*, pages 234–245. Springer, 2010.

[10] E. N. Mortensen and W. A. Barrett. Intelligent scissors for image composition. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 191–198. ACM, 1995.

[11] M. Tang, L. Gorelick, O. Veksler, and Y. Boykov. Grabcut in one cut. In *ICCV*, pages 1769–1776. IEEE, 2013.

[12] N. Xu, N. Ahuja, and R. Bansal. Object segmentation using graph cuts based active contours. *CVIU*, 107(3):210 – 224, 2007.