TOUCHCUT: SINGLE-TOUCH OBJECT SEGMENTATION DRIVEN BY LEVEL SET METHODS

Tinghuai Wang[⋆†] *Bo Han*[⋆] *John Collomosse*[†]

[⋆] Sony China Research Laboratory, Beijing, China
[†] CVSSP, University of Surrey, United Kingdom

ABSTRACT

In this paper, we propose an object segmentation algorithm driven by minimal user interactions. Compared to previous user-guided systems, our system can cut out the desired object in a given image with only a single finger touch minimizing user effort. The proposed model harnesses both edge and region based local information in an adaptive manner as well as geometric cues implied by the user-input to achieve fast and robust segmentation in a level set framework. We demonstrate the advantages of our method in terms of computational efficiency and accuracy comparing qualitatively and quantitatively with graph cut based techniques.

Index Terms— object segmentation, level set methods

1. INTRODUCTION

Object segmentation within natural images, i.e. extracting the foreground object (or object of interest) out of the cluttered background is an inherently ambiguous and challenging problem. In the absence of high level knowledge, there can be more than one interpretation of the foreground. Practical systems incorporate prior information via user interaction and low-level cues such as color and edges observed in the image.

A variety of interaction forms, ranging from roughly marking the desired boundary [1] to loosely drawing scribbles labeling the desired object and the background [2, 3], to placing a bounding box around the desired object [4], have been used. Regardless of the intervention modality, the goal of any interactive image segmentation is to minimize the amount of effort to cut out a desired object while accurately selecting objects of interest.

The popular graph cut based approaches [2, 4] balance the probability of pixels belonging to the foreground and background and the edge contrast. However, there is an inherent bias of graph cut towards shorter paths, as the boundary term sums over the boundary of the segmented regions. The level set based methods, on the contrary, include a length-based "ballooning" term which encourages a larger object segment. One application of level set methods to image segmentation has been the edge-based active contour model [5], which depends on image gradient and therefore is a rather local approach sensitive to noise. More robust approaches that encode the region information has been proposed later in [6, 7]. Higher level prior knowledge such as geometric shape priors has been introduced in [8].

The methods mentioned above inevitably experience interaction difficulties when adopted in small size touch screen applications, such as intelligent focus, white balance, and dynamic range functions in digital camera, digital camcorder or smart phones. In terms of convenience, a single finger touch is the most user-friendly interaction for object segmentation in these applications. In this paper, we introduce a novel object segmentation algorithm driven by minimum user interaction, i.e. a single touch on the image. The core contribution of this paper is an adaptive probabilistic edge-region-geometry description of the segmentation problem. By leveraging the flexibility of level set methods in energy minimization, the proposed method enables desired segmentation with accurate boundary placement and strong region connectivity while requiring minimum user interaction.

2. LEVEL SET REVISITED

The active contour models implemented via level set methods is a contour $\mathcal{C}$ in a domain Ω represented by the zero level set of a higher level embedding function ϕ: $\Omega \rightarrow \Re$. Evolving the contour $\mathcal{C}$ is achieved by evolving the level set function ϕ. The evolution of the level set function ϕ is governed by a partial differential equation (PDE). One can directly derive the PDE from a certain energy functional $E(\phi)$ on the space of level set functions and derive the Euler-Lagrange equation which minimizes $E(\phi)$: $\frac{\partial \phi}{\partial t} = -\frac{\partial E(\phi)}{\partial \phi}$.

3. SEGMENTATION FRAMEWORK

In the level set paradigm, we propose a new energy functional taking account of probabilistic edge map, color distribution of foreground and background in an adaptive manner as well as the geometric cue implied by user-input:

$$E(\phi) = E_e(\phi) + E_a(\phi) + E_b(\phi) + E_u(\phi) + E_s(\phi) + E_d(\phi) \quad (1)$$

where $E_e(\phi)$ is the edge probability term, $E_a(\phi)$ is the ballooning term, $E_b(\phi)$ is the Bayes statistical error term based on color distributions, $E_u(\phi)$ is the foreground consistency term, $E_s(\phi)$ is the geometric cue term, and $E_d(\phi)$ indicates the distance regularization term to ensure the stable evolution of the level set function by penalizing the deviation of the level set function from a signed distance function. Defining the distance regularization term is beyond the scope of this paper, readers are referred to [9] for details. These terms

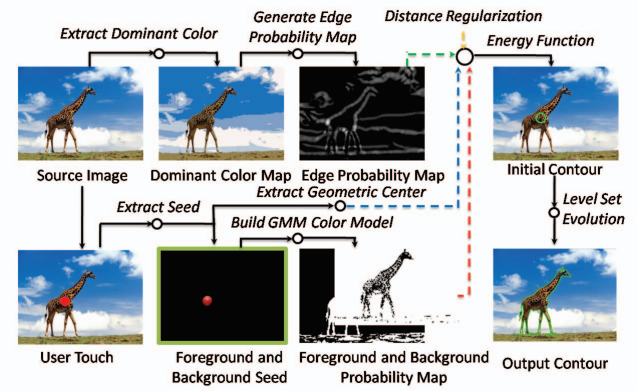

Fig. 1. System overview. Dominant color extraction is performed on the input image for calculating the edge probability map (first row). Foreground/background color model is estimated based on user input and the image border respectively (second row). The energy function incorporates the various energy terms. The evolution of the embedding function ϕ is specified by the energy function (right column). The zero level contour converges to the object boundary to generate the segmentation (bottom right).

can be categorized as: edge based energy, statistical prior energy, geometry energy and distance regularization energy. Fig. 1 presents an overview of the proposed system, where the dashed lines indicate these four energy categories. Each individual energy term is detailed in the following subsections.

3.1. Edge Based Energy

Classical snakes and active contour models [5] typically use an edge detector to halt the evolution of the curve on the boundary of the desired object. The gradient based edge detector inherently captures high frequency information but not necessarily the real boundary of the desired object. Moreover, it is also sensitive to noise. The edge-based active contour model is thus not applicable to most natural images especially texture rich or noisy data.

In order to describe the edge probability of color-texture homogeneous region in natural image, we adopt a similar approach to JSEG [10], which calculates an edge indicator J by observing the local distribution of color class labels without estimating a specific model for a texture region. In our proposed method, the color class labels are generated by extracting the dominant color (DC) modes and assigning each pixel with the label of according DC. The value of J is large near the boundaries of color-texture homogeneous region and small in region interiors, and thus can serve as edge "probability" while suppressing noise in texture region.

We propose a non-parametric DC extraction algorithm which considers both color distribution and color similarity to better explore the inherent characteristics. In the algorithm, colors in the CIE L*a*b* histogram are first clustered via a watershed-like process. Considering the peaks in the 3D histogram as islets in a lake, some of them are merged as the water level in the lake decreases, to make the algorithm robust to the noise (roughness) in the histogram. Finally, each cluster corresponding to a large enough proportion of pixels is extracted as a DC. An example of the DC extraction result is shown in the second image of the first row in Fig. 1.

E_e incorporates the edge indicator J and is defined similarly as the geodesic model [5] $E_e(\phi) = \mu_e \int_\Omega g\delta(\phi)|\nabla\phi|d\mathbf{x}$ where $g = \frac{1}{1+cJ}$, μ_e is the coefficient, c is a constant, H is the Heaviside function and δ is the Dirac delta function.

We define the ballooning term as $E_a(\phi) = \mu_b \int_\Omega gH(\phi)d\mathbf{x}$, which computes a weighted area of the region $\Omega^+_\phi \triangleq \{\mathbf{x} : \phi(\mathbf{x}) > 0\}$. This energy is introduced to speed up the motion of the zero level contour in the evolution process when the initial contour is not placed in the vicinity of the desired object boundary. The ballooning of the zero level contour is inhabited near the boundaries where J takes larger values.

3.2. Statistical Prior Energy

An optimal partition $\mathcal{P}(\Omega)$ of the image plane Ω can be computed by maximizing the *a posterior* probability $p(\mathcal{P}(\Omega)|I)$ for the given image I [11]. Applying Bayes' rule, it can be expressed as $p(\mathcal{P}(\Omega)|I) \propto p(I|\mathcal{P}(\Omega))p(\mathcal{P}(\Omega))$. $p(\mathcal{P}(\Omega))$ allows to introduce prior knowledge such as geometric priors to cope with missing low-level information. Under the given prior, optimal two-region partition is achieved by maximizing $p(I|\mathcal{P}(\Omega)) = p(I|\Omega^+)p(I|\Omega^-)$, where Ω^+ and Ω^- represent the regions inside and outside the contour respectively. Maximization of the *a posterior* probability is equivalent to minimizing its negative logarithm, we define $E_b(\phi)$ as

$$E_b(\phi) = -\mu_b[\log p(I|\Omega^+) + \log p(I|\Omega^-)]. \quad (2)$$

We assume that the image I in each region is characterized by the individual pixel values at different locations $\mathbf{x}$ and the pixel values are i.i.d. Let $\phi(\mathbf{x}) > 0$ if $\mathbf{x} \in \Omega^+$ and $\phi(\mathbf{x}) < 0$ if $\mathbf{x} \in \Omega^-$. We reduce (2) to $E_b(\phi) = -\mu_b \int_\Omega (H(\phi)\log p(I(\mathbf{x})|\theta^+) + (1-H(\phi))\log p(I(\mathbf{x})|\theta^-))d\mathbf{x}$. where θ^+ and θ^- represent the foreground and background color model respectively and μ_b is the coefficient which is specified in the implementation subsection.

The foreground and background color model are represented by Gaussian Mixture Model (GMM) learned from observations of pixels; specifically the pixels in the user-specified area are assumed to be foreground and the border of the image is assumed to be the background.

The user-specified area is usually a part of the desired object, and thus the foreground color model has higher confidence than the background color model, especially when the desired object intersects the border of the image. We propose a foreground consistency term to enforce the minimization of foreground statistical error as

$$E_u(\phi) = \frac{\mu_u \int_\Omega H(\phi)(1 - p(I(\mathbf{x})|\theta^+))d\mathbf{x}}{\int_\Omega H(\phi)d\mathbf{x}} \quad (3)$$

where μ_u is a coefficient specified in subsection 3.5.

3.3. Geometry Energy

People tend to select the geometrical centre when they are indicating the object of interest. Although not a precise mea-

surement, such a geometrical constraint provides a weak cue for the contour evolution process. We propose a central symmetry term to reflect this geometrical constraint, by computing the spatial deviation of the geometrical centre of zero level contour from the user-input point as

$$E_s(\phi) = \mu_s |\frac{\int_\Omega H(\phi)(\mathbf{x} - \overline{\mathbf{x}})d\mathbf{x}}{\int_\Omega H(\phi)d\mathbf{x}}| \quad (4)$$

where $\overline{\mathbf{x}}$ represents the user-input point. As the desired object could have very complex shape, this term is regarded as a relatively weak indication of the desired object's geometry.

3.4. Adaptive Weighting

Minimizing the proposed energy functional (1) with constant coefficients usually gives good segmentations. However, when the foreground and background distribution is not distinct, the Bayes error term would be non-discriminative and the contour evolution process would not converge to the desired object boundaries. In this case, the weight of Bayes' error term should be relatively small to increase the influence of other reliable terms. We expect it to be adaptively tuned based on the color modeling error on a per image basis. To this end, we estimate the misclassifying error in foreground/background seeds based on the posterior probability

$$\eta = \frac{1}{|\Omega^+|}\sum_{\mathbf{x}\in\Omega^+} p(I(\mathbf{x})|\theta^-) + \frac{1}{|\Omega^-|}\sum_{\mathbf{x}\in\Omega^-} p(I(\mathbf{x})|\theta^+)$$

and define coefficient $\mu_b = \max\{\overline{\mu_b}(1-\eta), 0\}$. When the misclassifying error η is close to zero, the weight approaches $\overline{\mu_b}$. When the color models are indistinct, μ_b approaches 0.

3.5. Gradient Descent Flow and Implementation

We use the standard gradient descent method to minimize the energy functional (1)

$$\frac{\partial\phi}{\partial t} = -\frac{\partial E_e(\phi)}{\partial\phi} - \frac{\partial E_b(\phi)}{\partial\phi} - \frac{\partial E_u(\phi)}{\partial\phi} - \frac{\partial E_s(\phi)}{\partial\phi} - \frac{\partial E_a(\phi)}{\partial\phi} - \frac{\partial E_d(\phi)}{\partial\phi}$$

where the gradient flows are deducted as follows:

$$\begin{aligned}
\frac{\partial E_e(\phi)}{\partial\phi} &= \mu_e\delta(\phi)\mathrm{div}(g\frac{\nabla\phi}{|\nabla\phi|}) \\
\frac{\partial E_b(\phi)}{\partial\phi} &= \mu_b\delta(\phi)\log\frac{p(I(\mathbf{x})|\theta^+)}{p(I(\mathbf{x})|\theta^-)} \\
\frac{\partial E_u(\phi)}{\partial\phi} &= \mu_u\delta(\phi)[\frac{(1-p(I(\mathbf{x})|\theta^+))}{(\int_\Omega H(\phi)d\mathbf{x})^2} \\
&\quad -\frac{\int_\Omega(1-p(I(\mathbf{x})|\theta^+))H(\phi)d\mathbf{x}}{(\int_\Omega H(\phi)d\mathbf{x})^2}] \\
\frac{\partial E_s(\phi)}{\partial\phi} &= \mu_s\delta(\phi)\frac{|(\mathbf{x}-\overline{\mathbf{x}}) - \int_\Omega(\mathbf{x}-\overline{\mathbf{x}})H(\phi)d\mathbf{x}|}{(\int_\Omega H(\phi)d\mathbf{x})^2}
\end{aligned}$$

$$\frac{\partial E_a(\phi)}{\partial\phi} = \mu_a g\delta(\phi) \quad \frac{\partial E_d(\phi)}{\partial\phi} = \mu_d\mathrm{div}(\frac{\mathcal{P}'(|\nabla\phi|)}{|\nabla\phi|}\nabla\phi)$$

In the implementation, the Heaviside function H is approximated by a smooth function defined by

$$H_\epsilon(x) = \begin{cases} \frac{1}{2}(1 + \frac{x}{\epsilon} + \frac{1}{\pi}\sin(\frac{\pi x}{\epsilon}), & |x| \le \epsilon \\ 1, & x > \epsilon \\ 0, & x < -\epsilon. \end{cases} \quad (5)$$

and the Dirac delta function δ is approximated by

$$\delta_\epsilon(x) = \begin{cases} \frac{1}{2\epsilon}(1 + \cos(\frac{\pi x}{\epsilon})), & |x| \le \epsilon \\ 0, & |x| > \epsilon. \end{cases} \quad (6)$$

We adopt the narrow band method [12] to substantially reduce the computational cost of level set method by confining the computation to a narrow band around the zero level set contour. In our prototype, we use a mouse click and a fixed brush size σ to simulate the user finger-touch. The embedding function ϕ is initialized by extracting the contour of the brush stroke. We empirically choose the parameters in the formulation as follows: $\mu_e = 6$, $\mu_b = 1$, $\mu_u = 10$, $\mu_s = 5$, $\mu_a = -3.8$, $\mu_d = 0.04$, $\sigma = 24$, $\epsilon = 1.5$ and 200 iterations of evolution.

4. RESULTS AND CONCLUSION

We have applied the proposed algorithm on a dataset consisting of 100 images from BSDS300 [13], GrabCut dataset [4] and the internet. We assess segmentation performance on both the qualitative and quantitative basis.

Fig. 2 presents the qualitative comparison of the proposed method with standard graph cut (middle) [2] and GrabCut (right) [4]. Graph cut approach is adapted such that the modeling of color distributions is exactly the same with the proposed approach to make a fair comparison with one single touch, i.e. the foreground is modeled from the pixels in user-touch area while the background is modeled by taking pixels from the border of the image. With significantly less user input, our method gives satisfactory segmentation even when the indistinct foreground and background color (first row) or complex topology (second row) present. We can see that graph cut approach fails to separate the objects exhibiting similar color with the desired object, whilst our approach fills the desired region by expanding from the interior of the selected object outwards and explicitly considers the object boundary and geometric property. GrabCut presents better spatial constraints than graph cut, benefiting from the bounding box while failed to exclude exotic regions (see the different levels of luminance underneath the tiger) which do not appear outside the bounding box, and also it suffers from the inherent short-cut problem (see the elephant's legs and nose).

For objective evaluation, we adopt the Berkeley Segmentation Benchmark [13] to evaluate segmentation against manual ground-truth. This benchmark considers two aspects of segmentation performance. Precision measures the fraction of true positives in the contours produced by a segmentation algorithm. Recall indicates the fraction of ground truth boundaries detected in the segmentation. The global F-measure, defined as the harmonic mean of precision and recall, provides a useful summary score for the segmentation algorithm [13]. Our proposed method receives a F-measure of 0.765 which outperforms the adapted graph cut (F-measure 0.538) and GrabCut (F-measure 0.697).

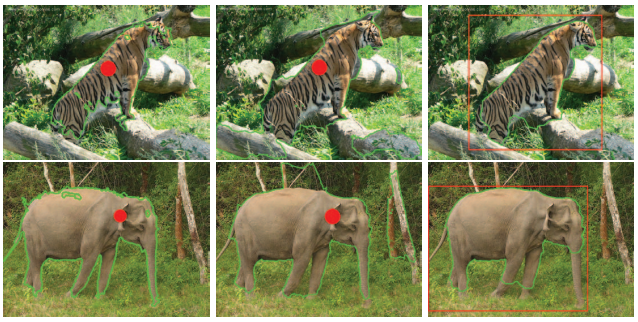

Fig. 2. Comparison of proposed method (left) with graph cut (middle) and GrabCut (right). The contour of segmented object is shown in green.

Fig. 3 presents some subjective qualitative segmentation results [1]. The first row shows the results on highly-textured images. The edge probability map enables the contour evolution over color-texture homogeneous regions without being stopped at local minimum. The second row shows the segmentation results of images with indistinct foreground and background colors. In this case, the color modeling error is large which adaptively results in a small weight on color based term E_b. On the other hand, the foreground consistency term E_u enforces the region inside the zero level contour to be coherent in the sense of color distribution regardless the background color distribution. Such a constraint significantly imposes the stability of the contour evolution process in the case of indistinct color distributions. The third row gives some segmentation results to deal with objects with complex shape. By leveraging the strength of the implicit contour representation in level set methods, our system is robust in coping with complex topologies without exhibiting short-cutting problem which is common in graph-cut based systems. The system is able to cope with weak boundaries and complex foreground and background, to extract meaningful object in most cases. The running time on a Core2 2.66 GHz PC is ~ 0.4 second per VGA image (640×480).

In summary, we presented a single-touch object segmentation system using level set methods. We demonstrated that by exploring the edge probability of color-texture homogeneous region as well as the statistical prior inferred from user input, our edge-region-geometry based model is able to robustly tackle the interactive object segmentation problem. By leveraging the flexibility of level set methods in energy minimization, our system achieved promising result in various natural images with complex scenes and objects. We believe that single-touch image manipulation will become increasingly important on emerging tablet form factor devices.

5. REFERENCES

[1] J. Wang, M. Agrawala, and M. F. Cohen, "Soft scissors: an interactive tool for realtime high quality matting," in *SIGGRAPH*. 2007, pp. 585–594, ACM.

[1] More results can be viewed online at: `http://personal.ee.surrey.ac.uk/Personal/Tinghuai.Wang/TouchCut`

Fig. 3. Representative segmentation results from our dataset.

[2] Yuri B. and Marie pierre J., "Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images," in *ICCV*, 2001, pp. 105–112.

[3] A. Protiere and G. Sapiro, "Interactive image segmentation via adaptive weighted distances," *IEEE Trans. Image Processing*, vol. 16, 2007.

[4] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut - interactive foreground extraction using iterated graph cuts," in *SIGGRAPH*. 2004, ACM.

[5] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," in *ICCV*, 1995, pp. 694–699.

[6] N. Paragios and R. Deriche, "A pde-based level-set approach for detection and tracking of moving objects," in *ICCV*, 1998, pp. 1139–1145.

[7] T. Chan and L. Vese, "Active contours without edges," *IEEE Trans. Image Processing*, pp. 266–277, 2001.

[8] D. Cremers, F. R. Schmidt, and F. Barthel, "Shape priors in variational image segmentation: Convexity, lipschitz continuity and globally optimal solutions," in *CVPR*, 2008, pp. 1–6.

[9] C. Li, C. Xu, C. Gui, and M. D. Fox, "Level set evolution without re-initialization: A new variational formulation," in *CVPR*. 2005, pp. 430–436, IEEE.

[10] B. S. Manjunath and Yining Deng, "Unsupervised segmentation of color-texture regions in images and video," *IEEE TPAMI*, pp. 1139–1145, 2001.

[11] N. Paragios and R. Deriche, "Geodesic active regions: a new paradigm to deal with frame partition problems in computer vision," *Journal of Visual Communication and Image Representation*, pp. 249–268, 2002.

[12] D. Adalsteinsson and J. Sethian, "A fast level set method for propagating interfaces," *Journal of Computational Physics*, pp. 269–277, 1995.

[13] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *ICCV*, 2001, pp. 416–423.