Saliency Cuts: An Automatic Approach to Object Segmentation

Yu Fu, Jian Cheng, Zhenglong Li and Hanqing Lu
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
P.O. Box 2728, Beijing 100190, P.R. China
{yfu, jcheng, zlli, luhq}@nlpr.ia.ac.cn

Abstract

Interactive graph cuts are widely used in object segmentation but with some disadvantages: 1) Manual interactions may cause inaccurate or even incorrect segmentation results and involve more interactions especially for novices. 2) In some situations, the manual interactions are infeasible. To overcome these disadvantages, we propose a novel approach, namely Saliency Cuts, to segment object from background automatically. By exploring the effects of labels to graph cuts, the so called “Professional Labels” is introduced to evaluate labels. With the aid of saliency detection, a multi-resolution framework is designed to provide “Professional Labels” automatically and implement object segmentation using graph cuts. The experiments demonstrate the promising performance of Saliency Cuts.

1. Introduction

Object segmentation is a challenging problem in computer vision, it is also a crucial task in many applications, such as object recognition, image classification, image editing and retrieval. During the last two decades, a lot of object segmentation methods have been developed, and these methods can be roughly classified into three categories. 1) Supervised methods: they need some kinds of prior information from training data which extremely limit the scope of applications in practise [1]. 2) Unsupervised methods: they are usually based on feature clustering which results in segmentation of similar patches [9] or aim to segment the common parts from a set of images [7]. 3) Semi-supervised methods: represented by interactive graph cuts [3, 8], which have witnessed an explosion of interest for their computation efficiency and effective performance.

For interactive graph cuts, object segmentation can be considered as a minimal graph cuts problem. First, users are required to label some regions as a prior, i.e. seeds. Then, a graph with two terminals is constructed and a so-called “Maxflow” optimization method [2] is applied to find the minimal cuts that corresponds to the desired segmentation, and this result can be re-corrected with continuously interactions. [3] obtained seeds from a few strokes or lines marked as object and background respectively by users. However, this manner may have some disadvantages. 1) Segmentation results heavily depend on the seeds selection which is a very skillful task. A novice often fails to provide effective seeds and more interactions are required for re-correcting. 2) In some situations, manual interactions are infeasible even only labeling a little portion. [8] focused on the first disadvantage by allowing users dragging a rectangle surrounding the desired object and getting seeds from it, but still involved manual interactions.

In recent years, the salience detection technique aiming at extracting the most possible attention concentrated regions in human vision theory are explored. Itti et al. [5] proposed a saliency model that simulated the visual search process of human; Liu et al. [6] designed a learning based saliency detection by feature extraction; Hou et al. [4] proposed a spectral residual approach to extract saliency map from the statistical singularities of an image’s log spectrum. These saliency regions are largely related to object regions and can be helpful to provide seeds for graph cuts implementation.

In this paper, we propose an automatic object segmentation approach integrating saliency detection and graph cuts, namely Saliency Cuts, to overcome the disadvantages of interactive graph cuts. We adopt the spectral residual approach [4] as the saliency detection scheme for its low computational cost and unsupervised manner. We also explored the effects of labels to graph based segmentation, and the so called “Professional Labels” is introduced to evaluate labels and a multi-resolution framework is designed to provide such “Professional Labels” automatically. In our
method, first, a coarse segmentation based on labeling of saliency detection is computed in low resolution level. Then, “Professional Labels” are constructed in high resolution level with the aid of this coarse segmentation. In the end, graph cuts are implemented and satisfying results are obtained.

The proposed Saliency Cuts have the following advantages:

1. **Professional labeling**: effective labeling seeds are provided, from which, complete object segmentation can be achieved.

2. **Automatic implementation**: none of the manual interactions are required.

The rest of this paper is organized as follows. In Sec. 2, we present the proposed Saliency Cuts, experimental results are reported in Sec. 3 and a conclusion follows in Sec. 4.

2. Saliency Cuts

In this section, Saliency Cuts are described in detail. First, we will briefly review the graph cuts based object segmentation [3].

A graph can be represented by $G = (V, E)$ where $V$ is the set of all nodes and $E$ is the set of all edges connecting nodes. Object segmentation can be considered as a minimal graph cuts problem: the nodes in the graph are represented by pixels except for two terminals $\{s, t\}$ denoting “object” and “background” respectively; the edges connecting nodes to terminals are $t - \text{links}$ and the edges connecting neighborhood nodes are $n - \text{links}$. Let $i \in V, x_i = 1$ or $0$ if $i$ is in object or background respectively, the target optimization function for graph cuts is:

$$E(X) = \sum_{i \in V} E_1(x_i) + \lambda \sum_{(i,j) \in E} E_2(x_i, x_j),$$

(1)

where $E_1(x_i)$ is the region energy corresponding to $t - \text{links}$, denoting the cost when the label of node $i$ is $x_i$, $E_2(x_i, x_j)$ is the boundary energy corresponding to $n - \text{links}$, denoting the cost when the labels of adjacent nodes $i$ and $j$ are $x_i$ and $x_j$ respectively, “object” and “background” feature distribution models based on seeds are constructed to encode $E_1$ and neighborhood similarities are used to encode $E_2$. The minimal cuts problem can be effectively solved by “MaxFlow”[2], and a corresponding object segmentation is obtained.

The label seeds play an important role in graph cuts based object segmentation. In the following subsection, we will demonstrate how does our method automatically and effectively provide seeds.

2.1. Automatic Labeling

**What are “Professional Labels”?** We have found in experiments that segmentation results heavily depend on label seeds for interactive graph cuts [3]: Effective labels can lead to complete object segmentation and reduce or eliminate the interactive efforts while novice labels usually result in region lost or false regions and require more manual interactions for re-correcting, we will demonstrate that in Sec. 3. As labeling is very skillful, we use “Professional Labels” to represent such effective labels. “Professional Labels” can be described as the labels that comprehensively cover the represented object and background regions and provide sufficient and accurate information to encode region energy. In interactive graph cuts [3], “Professional Labels” can be provided by skilled humans manually, but in our method, such labels are given by the multi-resolution framework automatically.

Given an image $I$, a low resolution image $I_1$ is obtained, assume the object labels and background labels in $I_1$ and $I$ are $L^o_1, L^b_1$ and $L^o, L^b$ respectively. Saliency detection is implemented and used as the coarse labels for $I_1$. The saliency map $[4]$ can be represented by:

$$S(x) = g(x) \ast \mathcal{F}^{-1}[\exp(R(f) + P(f))],$$

(2)

where $g(x)$ is a Gaussian filter; $\mathcal{F}^{-1}$ denotes the Inverse Fourier Transform; $R(f)$ is the spectral residual: $R(f) = L(f) - A(f)$, $L(f)$ represents the log spectrum of an input image, $A(f)$ is the general shape of log spectrum; $P(f)$ denotes the phase spectrum of the image. An object map $S^o(x)$ is formed by a binarization processing of $S(x)$.

Considering the general characteristics that interested objects are mostly concentrated at the center of images and set from the top bottom, some of the background regions are expected to be at the left and right side of images, and saliency areas here are usually noise. We adopt the object map areas within a proportion of photo size as our object labels $L^o_1(x)$, and set background labels $L^b_1(x)$ at the left and right side and bypass noisy saliency areas:

$$L^o_1(x) = \{x : S^o(x) = 1, r(x) < k_1w_1\},$$

(3)

$$L^b_1(x) = \{x : S^o(x) = 0, k_2w_1 < r(x) < \frac{w_1}{2}\}$$

where $r(x)$ is the horizontal distance to photo center, $w_1$ is the width of the low resolution image, and $k_1$ and $k_2$ are proportion coefficients.

The object labels in $I_1$ by saliency detection are usually isolated points or small areas centralized around the object boundary, we implement graph cuts and only get
a coarse object segmentation $C_1(x)$ with inexact boundaries. However, with the aid of $C_1(x)$, “Professional Labels” can be formed for $I$: let the coarse object region in $I$ derived from $C_1(x)$ be $O_1$. We shrink the boundary of $O_1(x)$ to avoid inexact boundaries and form a more accurate object labels $L^o$; expand it to form a ring region, accomplished background labels $L^b$ are the summation of the ring region and $L^b_1$:

$$L^o(x) = \{ x \in \varepsilon_{r_1}(O_1) \},$$

$$L^b(x) = \{ x \in (\rho_{r_2}(O_1) - \rho_{r_3}(O_1)) \cup L^b_1 \}.$$  \hspace{1cm} (4)

where $r_1 = l_1R(O_1)$, $r_2 = l_2R(O_1)$, $r_3 = l_3R(O_1)$. $R(O_1)$ is the minimal value of the width and length of the circumscribed rectangular of $O_1$, and $l_1$, $l_2$, and $l_3$ are proportion coefficients; $\varepsilon_{r_1}(O_1)$ is an erosion operator indicating shrinking region $O_1$ for $r_1$ pixels; $\rho_{r_2}(O_1)$ and $\rho_{r_3}(O_1)$ are the expansion operators denoting expanding region $O_1$ for $r_2$ and $r_3$ pixels respectively.

![Original image](a) Saliency map (b) Auto-labeling

**Figure 1. Automatic Professional Labels**

The labels represented by $L^o(x)$ and $L^b(x)$ are “Professional Labels”. First, object labels cover a large portion of the object and background labels are distributed in the image side and around the object which are sufficient to represent the background far from and near to object. One example is shown in Fig. 1, of which, (a) is original image, (b) is corresponding low resolution saliency map and (c) is the automatic labels obtained by our method: the blue regions are object seeds and the red regions are background seeds. Once the labeling of the image is finished, two sets of pixels labeled as “object” and “background” are defined as object seeds $O$ and background seeds $B$ respectively, the rest of unlabeled pixels are represented by $U$.

2.2. Graph Model Construction

Boundary energy $E_2$ is defined as follows:

$$E_2(x_i, x_j) = \exp \left( -\frac{(C_i - C_j)^2}{2\sigma^2} \right) \frac{1}{\text{dist}(i, j)}.$$ \hspace{1cm} (5)

where $(C_i - C_j)^2$ is the L2-Norm of the RGB color difference of two pixels $i$ and $j$, and $\text{dist}(i, j)$ is the position distance of two pixels $i$ and $j$.

Similar to [8], $K$ components GMMs for object seeds and background seeds represented by $G^o(x; \pi^o, \mu^o, \Sigma^o)$ and $G^b(x; \pi^b, \mu^b, \Sigma^b)$ respectively are used to encode region energy $E_1$:

$$E_1(x_k = 1) = 0 \quad E_1(x_k = 0) = C \quad \forall i \in O,$$

$$E_1(x_k = 1) = D^o(x_k) \quad E_1(x_k = 0) = D^b(x_k) \quad \forall i \in U,$$

where $D^o(x_k) = -\log(G^o(x_k; \pi^o, \mu^o, \Sigma^o))$, $D^b(x_k) = -\log(G^b(x_k; \pi^b, \mu^b, \Sigma^b))$, $C$ is a large constant to make sure the labeled pixel is assigned to the label value:

$$C = \max_j \sum_{j, (i,j) \in E} E_2(x_i, x_j)$$

A graph is constructed and we also adopt “MaxFlow” [2] to solve the minimizing cuts problem and obtain a corresponding segmentation.

3. Experiments

In order to verify the proposed method, we conduct extensive object segmentation on image database [6]. We also compare the Saliency Cuts with interactive graph cuts. For those approaches, GMMs of $K = 5$ components are applied. In Eq. (1), $\lambda = 50$ is adopted; in Eq. (3): $k_1 = 0.3$, $k_2 = 0.8$, $k_3 = 0.3$ are set; in Eq. (4): $l_1 = 0.18$, $l_2 = 0.07$, $l_3 = 0.2$ are adopted. These parameters are not unique, actually, our method works well in a range of parameters.

Segmentation of interactive graph cuts using manual strokes are presented in Fig. 2: (b), (c), (d), of which, manual labels of object (blue stroke region) and background (red stroke region) are in the top row and corresponding segmentation results are in the bottom row. Manual “Professional Labels” in (b) lead to a complete object segmentation, while novice labels fail to represent the object and background features and lead to bad segmentation results: the insufficient object seeds in (c) result in face region lost and insufficient background seeds in (d) bring in false object regions. Segmentation of GrabCuts [8] are present in Fig. 2(e). Automatic segmentation obtained by Saliency Cuts are present in Fig. 2(a): quite a complete object segmentation is obtained comparable to interactive graph cuts with manual “Professional Labels” in (b) and GrabCuts in (e).

We also test our methods on three typical kinds of images with the objects of humans, animals and plants. Segmentation results are partly presented in Fig. 3. We have to mention that our method mainly emphasize on the segmentation of single object and may fail when object region extend to the left and right side of the image as we take these region as background seeds.
4. Conclusions

We propose an novel approach namely Saliency Cuts to segment object automatically. With the aid of saliency detection, a multi-resolution framework is designed to provide “Professional Labels” as well as implement graph cuts. Experiments show that Saliency Cuts obtain quite complete object segmentation comparable to interactive graph cuts with manual ”Professional Labels” and GrabCuts, while our method is absolutely automatic and involves none of manual interactions. In the future work, we will relax the assumption for background regions’ distribution to promote the expansion of Saliency Cuts.

Acknowledgments. This work was supported by the National Natural Science Foundation of China, grant No.60605004 and No. 60723005.

References


