Journal of Visual Communication and Image Representation 55 (2018) 393–403

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

journal homepage: www.elsevier.com/locate/jvci

Toward a unified scheme for fast interactive segmentation☆

Ding-Jie Chen, Hwann-Tzong Chen*, Long-Wen Chang

Department of Computer Science, National Tsing Hua University, Hsinchu 300, Taiwan

ARTICLE INFO

Keywords:
Interactive
Image segmentation
Manifold ranking
Machine-assisted

ABSTRACT

This paper presents an efficient and effective interactive segmentation scheme for extracting the region of a foreground object in an image. Our goal is to design an interactive segmentation algorithm that unifies the bounding-box-based, seed-based, and query-based interaction mechanisms for pursuing (*i*) high efficiency in simple interaction mechanism, (*ii*) few interaction rounds, and (*iii*) short response time. The proposed algorithm starts with a user-provided bounding box and obtains candidate background superpixels for inferring the foreground object. Our algorithm tolerates imprecise bounding boxes and provides two kinds of interactions for acquiring correct labels from the user. The user can either input the seed/scribble annotations or label the algorithm-queried regions. Our algorithm selects the most uncertain region as a query, and this query-based interaction mechanism reduces the burden of the user on deciding suitable annotation locations. The average response time per-interaction of our algorithm is merely 0.014 s. Our experiments demonstrate that the algorithm achieves an efficient unified scheme for interactive image segmentation.

1. Introduction

Image segmentation is a fundamental problem in computer vision. Tasks of fully automated segmentation often suffer from ambiguities in the definition of the region of interest. On the other hand, tasks of fully manual segmentation are certainly time-consuming and by no means preferable. Hence, interactive image segmentation, *ie*, segmentation with a human in the loop, is more suitable for achieving satisfactory segmentation accuracy while keeping the time cost acceptable to the user [1–20]. Since the user inputs are required to guide the segmentation process, in order to get a smooth interaction experience, the response time of the segmentation algorithm is expected to be short enough, and the input mechanism is expected to be as simple as possible.

The existing interactive segmentation algorithms can be classified into four categories: (*i*) bounding-box-based [4,6,10,13,14,18], (*ii*) seed/scribble-based [3,5,7–9,11–13,16–19], (*iii*) contour-based [1,2,20], and (*iv*) query-based [15,21]. In general, it is easier for users to indicate the candidate foreground object via a bounding box. However, the segmentation accuracy is usually limited by how precisely the box is drawn. Both the seed-based algorithms and the contour-based algorithms can tackle the situations of complex-shaped objects as long as sufficient user inputs are given—more rounds of interactions are needed in comparison with the bounding-box-based algorithms. As to the query-based algorithm, though the user only needs to provide the true-or-false answers, sometimes the probability of hitting a foreground region is too low for some very small objects, which means that the number of interactions could be out of user's control to get a satisfactory segmentation result.

We propose a new interactive segmentation algorithm combining the advantages from different categories of interactive image segmentation algorithms. Our algorithm starts with a bounding box that roughly covers the foreground object. Then, our algorithm provides an user-assistance mechanism for subsequent improvements in the segmentation accuracy. The mechanism actively suggests the most uncertain region as a query for the user to respond with the correct binary label, in which the user only needs to give a true-or-false answer. An overview of the proposed algorithm is shown in Fig. 1. In this unified scheme, the initial user input, which is the bounding box, is used to provide the underlying distribution of background labels. Then, we estimate an initial confidence level of foreground object according to the potential background labels and the pairwise superpixel relevance using a manifold ranking strategy. Our user-assistance mechanism then acquires some label information from the user to update the segmentation result and the confidence level of the foreground object.

The contributions of the proposed unified scheme for interactive image segmentation are:

1. The initial estimation of the foreground object confidence is robust. We can allow part of the foreground object to straddle the bounding

☆ This paper has been recommended for acceptance by Radu Serban Jasinschi.

* Corresponding author.
E-mail address: htchen@cs.nthu.edu.tw (H.-T. Chen).

https://doi.org/10.1016/j.jvcir.2018.06.011
Received 15 July 2017; Received in revised form 12 April 2018; Accepted 13 June 2018
Available online 19 June 2018
1047-3203/

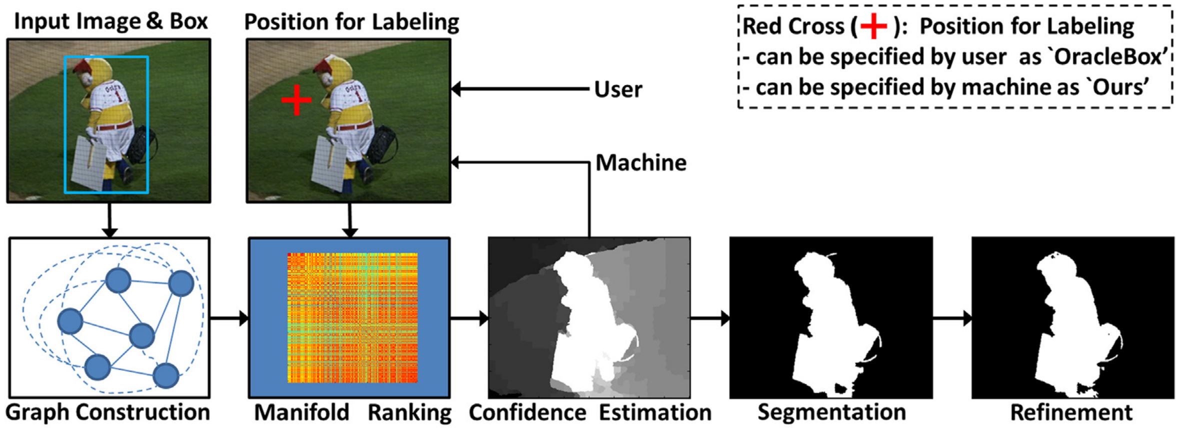

Fig. 1. An overview of the proposed interactive image segmentation algorithm. In 'BnS' approach, the user has to manually explore image positions for labeling. In 'BnQ' approach, the user only needs to provide the binary responses for labeling the automatically generated positions.

box. The user may casually draw a box that is not necessarily "bounding".

2. The response time is very short. The segmentation updating process is simple and efficient, which means that our algorithm is practical to be used in real-time incremental manipulation.
3. The input mechanism is flexible. After providing a bounding box, the user may manually input some labeled seeds to acquire a more accurate segmentation (which is called '*BnS*' as the abbreviation of 'Box and Seeds'), or the user gives true-or-false feedback to the query-region for easily obtaining acceptable segmentation results (which is called '*BnQ*' as the abbreviation of 'Box and Queries').

2. Related work

We classify previous interactive image segmentation algorithms into four categories, depending on the types of user inputs.

2.1. Bounding-box-based

Algorithms of this kind assume that the user specified bounding box encloses the foreground object. Then, figure-ground segmentation is performed according to this assumption. Rother et al. [6] propose the GrabCut algorithm that iteratively estimates the Gaussian mixture models of the foreground and the background areas [4], and then refines the segmentation using graph cuts. Lempitsky et al. [10] model the segmentation as an integer programming problem with the prior from the given bounding box. Tang et al. [13] propose a different energy term for global maximization in graph cut. It can also accept the seed-labels. Wu et al. [14] introduce a multiple-instance-learning algorithm to segment the foreground object inside a given bounding box. Zemene and Pelillo [18] assume that a foreground object is a dominant set [22], and introduce a constrained version of the dominant set algorithm that makes the generated dominant sets contain the user labeled regions.

2.2. Seed/scribble-based

Algorithms of this type solve image segmentation tasks according to user-labeled seeds. The segmentation results usually vary depending on the amount and the positions of the user annotated seeds. Boykov and Jolly [3] represent an image as a graph and treat the user inputs as hard constraints to find an optimal segmentation via graph cuts. Li et al. [5] propose the Lazy Snapping algorithm that contains an object marking step and a boundary editing step. Freedman and Zhang [7] use an object-specific shape prior to a graph cuts framework. Grady [8] models the segmentation result as the probability for each pixel first arrive the labeled pixels. Vicente et al. [9] use explicit connectivity prior to overcoming the shrinking bias effect of graph cuts. Gulshan et al. [11] use a star-convexity shape constraint to segment images under geodesic distance. Anh et al. [12] propose an algorithm to segment a set of multi-view images by interactively cutting a small image subset. Wang et al. [16] combine the region and boundary information to improve the conventional graph cut methods. Feng et al. [17] introduce the cue selection mechanism in a graph cuts framework, which based on the intuition that only one cue is needed at each vertex while optimizing the segmentation energy. Another bounding-box-based algorithm, proposed by Zemene and Pelillo [18], also accepts user scribbles to guide the constrained dominant set. Luo et al. [19] cast the segmentation problem as a multi-classification problem, and then learn a discriminative projection matrix through Fisher linear discriminant analysis for segmentation.

2.3. Contour-based

Algorithms in this category require the user to sketch the boundary of the foreground object roughly. Kass et al. [1] present a contour deformable algorithm to warp the user-specified contour for segmentation. Mortensen and Barrett [2] propose an intelligent scissors algorithm, which requires the user to put some seeds around the foreground object. Then the object contour is calculated via the shortest path for separating the figure and ground. Chan and Vese [20] propose an active-contour approach based on Mumford-Shah functional and level sets for segmenting objects whose boundaries have no need to be defined by the gradient.

2.4. Query-based

Algorithms in this category actively ask users the correct labels of uncertain regions for updating the segmentation. Rupprecht et al. [15] introduce an active region-querying image segmentation algorithm. The queried region is calculated using the geodesic distance transform and the Markov Chain Monte Carlo sampling. Chen et al. [21] propose an interactive segmentation algorithm which based on a transductive inference process. A query-pixel is selected by an entropy measurement on feature similarity and uncertainty [23].

2.5. Ours

This work illustrates a unified scheme for fast interactive image segmentation by integrating bounding-box-based, seed/scribble-based, and query-based mechanisms. Most of the existing methods [4,6,10,3,5,7–9,11–13,16–19] initialize a segmentation task with a