Neurocomputing 000 (2017) 1–17

Contents lists available at ScienceDirect

Neurocomputing

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

Spectral salient object detection

Keren Fu[a,*], Irene Yu-Hua Gu[b], Jie Yang[c]

[a] College of Computer Science, Sichuan University, Sichuan 610065, China
[b] Signal Processing Group, Department of Signals and Systems, Chalmers University of Technology, Gothenburg SE-41296, Sweden
[c] Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200240, China

ARTICLE INFO

Article history:
Received 14 May 2017
Revised 29 August 2017
Accepted 8 September 2017
Available online xxx

Communicated by Prof. Junwei Han

Keywords:
Salient object detection
Object holism
Normalized cut
Graph partition
Regional saliency

ABSTRACT

Many salient object detection methods first apply pre-segmentation on image to obtain over-segmented regions to facilitate subsequent saliency computation. However, these pre-segmentation methods often ignore the holistic issue of objects and could degrade object detection performance. This paper proposes a novel method, *spectral salient object detection*, that aims at maintaining objects holistically during pre-segmentation in order to provide more reliable feature extraction from a complete object region and to facilitate object-level saliency estimation. In the proposed method, a hierarchical spectral partition method based on the normalized graph cut (Ncut) is proposed for image segmentation phase in saliency detection, where a superpixel graph that captures the intrinsic color and edge information of an image is constructed and then hierarchically partitioned. In each hierarchy level, a region constituted by superpixels is evaluated by criteria based on figure-ground principles and statistical prior to obtain a regional saliency score. The coarse salient region is obtained by integrating multiple saliency maps from successive hierarchies. The final saliency map is derived by minimizing the graph-based semi-supervised learning energy function on the synthetic coarse saliency map. Despite the simple intuition of maintaining object holism, experimental results on 5 benchmark datasets including ASD, ECSSD, MSRA, PASCAL-S, DUT-OMRON demonstrate encouraging performance of the proposed method, along with the comparisons to 13 state-of-the-art methods. The proposed method is shown to be effective on emphasizing large/medium-sized salient objects uniformly due to the employment of Ncut. Besides, we conduct thorough analysis and evaluation on parameters and individual modules.

1. Introduction

Studies from neurobiology and cognitive psychology indicate that human brains are capable of selecting a certain visual contents for further processing [1,2]. Modeling human bottom-up visual attention on images, referred to as *bottom-up saliency detection*, is aimed at detecting salient image parts that can easily attract human attention. Bottom-up saliency detection has gained increasing research interest recently. Under this theme there are two sub-types [3,4], namely *eye fixation modeling* [4–7] and *salient object/region detection* [3,8,9]. In this paper, we address the second type. The recent advance in salient object detection is driven by high-level applications such as automatic object segmentation [10,11], content-aware image editing [12–15] and retrieval [16,17].

Many existing methods for salient object detection in still images, e.g., [9,18–21] essentially employ certain pre-segmentation techniques. Resultant regions/superpixels are then considered as basic processing units and fed into saliency computation. This mean not only facilitates computation but also avoid pixel-level noise. Typical techniques include clustering-based segmentation (e.g., Meanshift [22], SLIC superpixels [23]), or merging-based segmentation (e.g., graph-based [24]). Unfortunately, since these methods are based on local image properties, they could result in highly over-segmented regions, where an object breaks up into small regions that ignore the holistic object. Such regions not only easily introduce noise, but also keep one from assessing the object as a whole entity (Fig. 1). As a salient object detection method is aimed at emphasizing the entire object uniformly in the resultant saliency map [3] (the ground truth in Fig. 1), a pre-segmentation that is consistent to the human visual perception and retains holistic object, intuitively, can contribute to more accurate saliency estimation. Retaining holistic object allows more reliable feature extraction and analysis such as colors, shapes and texture from a complete object region.

Segmenting complete regions of arbitrary objects in a pre-segmentation stage is a very challenging task. This is because the task of saliency detection aims at detecting generic arbitrary

* Corresponding author.
E-mail address: fkrsuper@scu.edu.cn (K. Fu).

http://dx.doi.org/10.1016/j.neucom.2017.09.028
0925-2312/

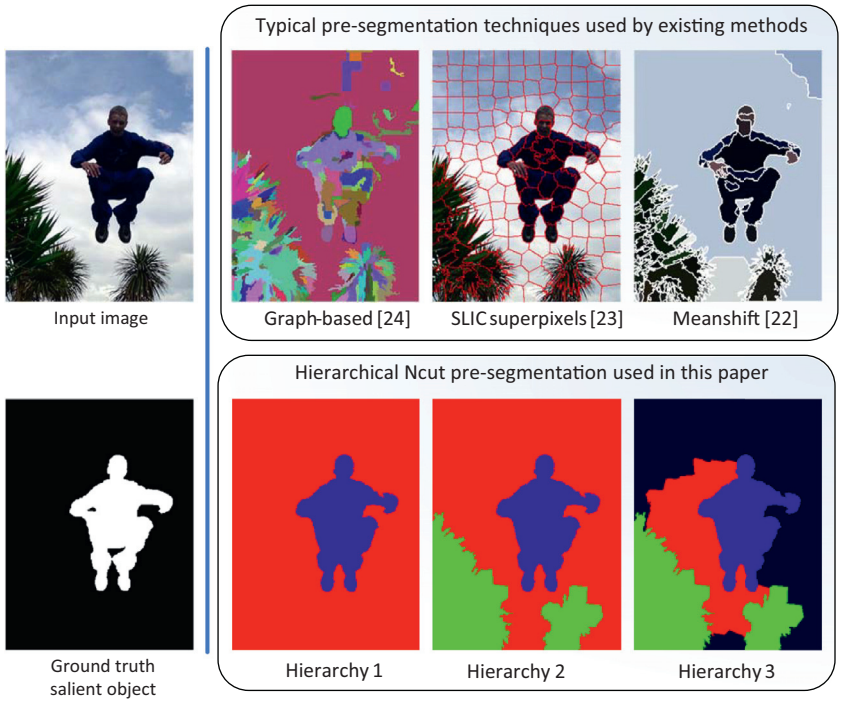

Fig. 1. Comparison between several commonly used segmentation methods and the proposed Ncut-based pre-segmentation. Row-1 (columns 2–4): graph-based segmentation [24], SLIC superpixels [23], and Meanshift segmentation [22]); Row-2 (columns 2–4): results from three hierarchies from the proposed method, where different segments are assigned with different colors. The parameters of the graph-based segmentation [24] were chosen in a similar way as in [9]. About 200 SLIC superpixels [23] were generated similar to [20,21]. The parameter setting of Meanshift [22] is similar to [8]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

objects which human eyes attend. Therefore, specifying category-dependent prior to assist segmentation is not feasible. To remedy this, we propose to utilize a spectral partition technique—the normalized graph cut (Ncut) [25] for salient object detection, because Ncut is unsupervised and shows more agreement to we human perception. As written by Shi and Malik [25], "Rather than focusing on local features and their consistencies in the image data, Ncut aims at extracting the global impression of an image". In this paper, we propose a hierarchical spectral partition method for the segmentation phase of saliency detection. The proposed method uses a superpixel graph to capture the intrinsic image color and edge information, and is based on the following observations: (a) Ncut has a strong discriminative power to separate image contents in object-level because it is a global criterion. It has hence the potential to maintain object holism and return complete boundary of an object (Fig. 1). (b) A salient object often has some unique appearance in terms of color or texture as comparing to its surroundings, implying some visual dissimilarity [5,9]. Therefore, a complete boundary of object is preferred by Ncut. A very low Ncut cost is often achieved if an entire object is separated from the remaining image.

Although Ncut has been widely used for image segmentation [25–27], applying it for inducing saliency computation and hence saliency maps is not well-studied. Furthermore, the aim of image segmentation is different from that of salient object detection, and hence, using Ncut solely is not adequate for rendering a saliency map. To further address this issue, we incorporate Ncut with several regional saliency metrics. Some other works [9,10,20,28,29] append graph cut to saliency maps as a second stage to achieve figure-ground segmentation. Our method differs from theirs since we firstly use Ncut to retain object holism prior to saliency computation, whereas [9,10,20,28,29] do not preserve object holism during saliency detection and might achieve less satisfactory saliency maps. In this paper, we believe conducting saliency detection on holism-retained segmentation could lead to better detection accuracy. Besides, the proposed method differs obviously from the above methods on technical aspects as well as implementations. The main novelties of this paper are summarized below:

1. For maintaining the holism of objects, Ncut is employed for separating salient image contents, where a novel hierarchical spectral partition method is introduced for pre-segmentation. It partitions a superpixel graph that captures the intrinsic color and edge information in an image. A binary segmentation tree is later generated, where an entire object is likely to be retained in emerging hierarchies.
2. For modeling image saliency in different hierarchies, Ncut is incorporated with regional saliency metrics. Three regional saliency metrics are introduced based on figure-ground principles and statistical prior. Salient objects are enhanced by integrating intermediate saliency maps from successive hierarchies.
3. Despite the simple intuition of maintaining object holism, we show that the proposed method achieves state-of-the-art performance on 5 benchmark datasets. Parameters of the proposed method are evaluated both quantitatively and comprehensively.

Although part of our work is published in the conference paper [30], this paper has significantly extended and improved our previous work, where we incorporate additional edge term in graph affinity computation, employ constrained Ncut for the first hierarchy for better cut initialization, and also conduct thorough evaluation on parameters and modules. In addition, more technical details and further extensive test results on objects in complex background are included.

The reminder of the paper is organized as follows. Section 2 describes the related work on salient object detection. Section 3 briefly reviews the fundamental of normalized graph cut, upon which our proposed method is built. Section 4 describes the proposed method in details. Experimental results,

Please cite this article as: K. Fu et al., Spectral salient object detection, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.09.028