ARTICLE IN PRESS

J. Vis. Commun. Image R. xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

High-level background prior based salient object detection ☆

Gang Wang, Yongdong Zhang *, Jintao Li

Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, China

ARTICLE INFO

Article history:
Received 17 August 2016
Revised 4 January 2017
Accepted 9 February 2017
Available online xxxx

Keywords:
Salient object detection
Background prior
Superpixel
Objectness

ABSTRACT

Salient object detection is a fundamental problem in computer vision. Existing methods using only low-level features failed to uniformly highlight the salient object regions. In order to combine high-level saliency priors and low-level appearance cues, we propose a novel Background Prior based Salient detection method (BPS) for high-quality salient object detection.

Different from other background prior based methods, a background estimation is added before performing saliency detection. We utilize the distribution of bounding boxes generated by a generic object proposal method to obtain background information. Three background priors are mainly considered to model the saliency, namely *background connectivity prior*, *background contrast prior* and *spatial distribution prior*, allowing the proposed method to highlight the salient object as a whole and suppress background clutters.

Experiments conducted on two benchmark datasets validate that our method outperforms 11 state-of-the-art methods, while being more efficient than most leading methods.

1. Introduction

In cognitive process, human can rapidly grasp the interesting and outstanding parts in a scene. To make the computer obtain this capability, increasing research efforts have been conducted for visual saliency analysis. Rather than eye fixation prediction just predicting a few salient points in an image, salient object detection aims to compute a salient map that highlights the most salient object regions, so it can not only speed up the following visual processes, but also improve the performance of many vision tasks [1,2].

Most of state-of-art methods are built on the assumption that salient object regions should have high appearance contrast from the background in an image, namely *contrast prior*. Local contrast based methods tend to find edges and corners, failing to highlight the whole object regions (Fig. 1(c)), while global contrast based methods sometimes highlight insignificant regions due to false judgements on object regions (Fig. 1(e)–(h)). Due to the complexity and diversity of salient object detection problem, modeling the foreground objects is hard and methods using purely low-level cures are not likely to generate high-quality results.

Instead of directly modeling the complex foreground objects, we tackle this problem using high-level background priors. Existing background prior based methods simply regard all the regions along the image boundary as background. This assumption is fragile and easily leads to false detection when the object touches the image boundary. What's more, we get the lessons from the previous methods that combined priors can improve the detection results. Besides the *background contrast prior*, two more priors are exploited to detect background regions, typically *background connectivity prior* and *spatial distribution prior*. The first prior assumes that the saliency of a pixel can be inferred by the appearance connectivity to the image border. To leverage this prior, the background connectivity salient map is measured by improved FastMBD [3]. The second prior assumes that a region is considered as background when it distributes vastly in an image. We measure the *spatial distribution prior* according to a fast online cluster results.

The main contributions of this paper are twofold:

1. Applying the high-level prior of objectness, we deal with the background estimation according to the distribution of bounding box proposals, which bridges the gap between object proposal problem and salient object detection problem at some extent.
2. With the help of three background priors, a combined method BPS is proposed for salient object detection, and achieves state-of-art performance while still being much efficient.

☆ This paper has been recommended for acceptance by Anan Liu.

* Corresponding author.
E-mail addresses: wanggang01@ict.ac.cn (G. Wang), zhyd@ict.ac.cn (Y. Zhang), jtli@ict.ac.cn (J. Li).

http://dx.doi.org/10.1016/j.jvcir.2017.02.004
1047-3203/© 2017 Elsevier Inc. All rights reserved.

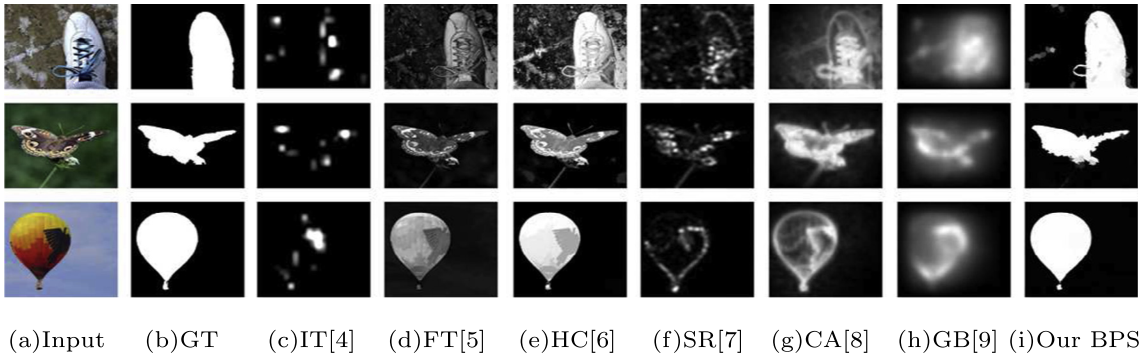

Fig. 1. Saliency maps generated by *contrast prior* based methods (c-h), and by our proposed BPS method. GT is the ground truth, and other methods can be inferred from the references. (See above-mentioned reference for further information.)

In addition, the visual unit in our algorithm is superpixel, however the superpixel segmentation is still a speed bottle neck as a pre-processing step. For practical reason, we optimize the SLIC [10] algorithm with AVX instruction set. The optimized algorithm runs about 30 ms per image, more than 6× speed up compared to the original algorithm (200 ms).

2. Related work

Visual saliency analysis is a highly active research direction in computer vision community because many other tasks [11–13] can benefit from it. This task has been extensive studied by plenty of works, which can be divided into two major categories: eye fixation prediction and salient object detection. This paper mainly focuses on the latter issue. Thorough discussion about this field is beyond the scope of this paper, we infer interest readers to a comprehensive survey [14] for more information.

A considerable amount of visual saliency models gained inspirations from the research conclusions of cognitive psychology [15]. Early stage methods mainly depended on local *contrast prior*, and it assumed that a salient pixel/region should present high contrast within a certain scope neighborhood. Itti et al. [4] proposed the first well-known local computational method, which measured the saliency map according to the center-surrounded contrast using multi-scale Difference of Gaussian. A lot of following methods were influenced by this work using other contrast information, including edge contrast [16], curvature [17] and self-similar difference [18]. The weakness of local contrast based methods is obvious: the saliency maps generated by these methods tend to find contours of object well, however the object interior is attenuated. Global contrast based methods measure the saliency map based on entire image information. Hou and Zhang [7] defined image saliency using frequency domain processing. Using different sizes of filters, their work had the capabilities to find different scales of salient object regions, but the result was rather fuzzy. An efficient region contrast based work [19] achieved significantly improved performance using color histogram based contrast. Their work had the advantage of assigning equal saliency to similar appearance regions, but still had difficulty in highlighting the entire object regions. Global methods alleviate the limitations of local methods at some extent, but the problem of "object attenuation" still exists. The essential problem is that the salient object detection is actually a binary segmentation issue, while using purely *contrast prior* is unlikely to separate the salient object from background totally.

Instead of directly dealing with complicate foreground detection, many methods tackled the salient object detection problem in an opposite way. Another two priors, namely *background connectivity prior* and *spatial distribution prior*, were used to model the background regions. *Background connectivity prior* assumes that most background regions can be easily connected to image boundaries, while this is much harder for object regions. Based on this important observation, both Wei et al. [20] and Qin et al. [21] computed the saliency using geodesic distance. Later, Zhang et al. [3] proposed an efficient method called FastMBD, which achieved promising performance. This strategy is also used in our work, but with the following two differences: (1). Rather than simply regarding the image boundary as background region, we pick up the representative colors to describe the background information, according to the pooling results of bounding boxes generated by the general object proposal algorithm BING [22]. (2). Undesirable distance accumulation happens with the path heading to the image center, so we take a simple but effective method to clip the accumulation.

Spatial distribution prior assumes that the foreground region often locates within a certain part of the image, while the background region is close to the image border and tends to spread across the image. This prior was employed by Gopalakrishnan et al. [23] and Ye et al. [24]. They both needed to select a threshold to binarize the saliency map. However, the optimal threshold was hard to determine, which may hurt the performance. Instead of relying on the threshold selection, we measure the spatial variance according to a fast cluster algorithm [25].

In recent years, convolutional neural network (CNN) has been successfully applied to many computer vision tasks, including salient object detection. Li and Yu [26] proposed a visual saliency method by predicting a saliency score for each superpixel via deep features. Liu and Han [27] tackled the salient object detection problem in a coarse to fine manner. They first generated rough saliency map through CNN, and hierarchically and progressively improved the image details by combining the global and local information. Kuen et al. [28] also employed similar strategy, but they refined the details according to a recurrent attentional network. Their work had better ability to deal with flexibly-sized objects in the image. Though deep learning (DL) based methods are good at evaluating objectness by extracting high-level features, they are relatively weak in sharping the object boundary. In addition, DL is difficult to learn some prior features, such as background connectivity, compactness, which are proved to be efficient in hand-designed methods. What is more, advanced modeling and clustering methods are very promising for this task. Liu et al. originally proposed the hierarchical clustering multi-task learning algorithm for joint model learning and grouping and explored the latent relationship among multiple clustering [29]. The discovered clustering information and the identified category information can further benefit the salient region segmentation.

Please cite this article in press as: G. Wang et al., High-level background prior based salient object detection, J. Vis. Commun. (2017), http://dx.doi.org/10.1016/j.jvcir.2017.02.004