Biologically Inspired Cognitive Architectures (2014) 9, 1–8

Available at www.sciencedirect.com
ScienceDirect
journal homepage: www.elsevier.com/locate/bica

LETTER

A salient object detection framework beyond top-down and bottom-up mechanism

Duzhen Zhang [a,b,*], Chuancai Liu [a]

[a] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu, PR China
[b] School of Computer Science and Technology, Jiangsu Normal University, Xuzhou 221116, Jiangsu, PR China

Received 2 March 2014; received in revised form 24 June 2014; accepted 26 June 2014

KEYWORDS
Salient object;
Bottom-up;
Top-down;
Selection history;
Physical salience;
Current goal

Abstract
Traditional saliency-based attention theory supposed that bottom-up and top-down factors combine to direct attentional behavior. This dichotomy fails to explain a growing number of cases in which neither bottom-up nor top-down can account for strong selection biases. Thus, the top-down versus bottom-up dichotomy is an inadequate taxonomy of attentional control. In this study, a general computational salient objects detection framework beyond top-down and bottom-up mechanism is presented. It possesses three parts: selection history, current goal and physical salience. Selection history is integrated with current goal and physical salience to compose an integrative framework. An image window saliency is defined as the objectness score of the window. Experimental results on challenging object detection datasets demonstrate that physical salience generates bottom-up saliency map for highlighting the salient regions of image, the main effect of selection history is to concentrate on salient objects, the current goal has strong effect to detect correct salient objects.

Introduction

Visual search plays a key role in our everyday activities; the visual system pays attention to the salient objects for efficient search. Visual saliency plays important roles in natural vision in that saliency can direct eye movements, deploy attention, and facilitate tasks like object detection and scene understanding. Many models have been built to compute saliency maps. There are two conventional categories of factors that drive attention: bottom-up and top-down factors (Desimone & Duncan, 1995). Bottom-up factors are derived solely from the visual scene. Regions of interest that attract our attention are in a bottom-up way and the

* Corresponding author at: School of Computer Science and Technology, Jiangsu Normal University, Xuzhou 221116, Jiangsu, PR China.
E-mail address: zhduzhen@aliyun.com (D. Zhang).

http://dx.doi.org/10.1016/j.bica.2014.06.005
2212-683X/

responsible feature for this reaction must be sufficiently discriminative with respect to surrounding features. Inspired by the feature-integration theory (Treisman & Gelade, 1980), Itti, Koch, and Niebur (1998) proposed one of the earliest bottom-up selective attention models by utilizing color, intensity and orientation of images. Most computational models (Bruce & Tsotsos, 2005; Zhang, Tong, Marks, Shan, & Cottrell, 2008; Achanta, Estrada, Wils, & Süsstrunk, 2008; Achanta, Hemami, Estrada, & Süsstrunk, 2009; Harel, Koch, & Perona, 2007; Goferman, Zelnik-Manor, & Tal, 2010; Chang, Liu, Chen, & Lai, 2011; Cheng, Zhang, Mitra, Huang, & Hu, 2011; Gopalakrishnan, Hu, & Rajan, 2010; Klein & Frintrop, 2011; Lu, Zhang, Lu, & Xue, 2011; Perazzi, Krahenbuhl, Pritch, & Hornung, 2012) are data-driven and focused on bottom-up attention, where the subjects are free-viewing a scene and salient objects attract attention. Bottom-up attention can be biased toward targets of interest by top-down cues such as object features, priors, reward, scene context and task demands. Top-down methods (Liu et al., 2011; Yang & Yang, 2012) are task-driven or goal-driven. This entails supervised learning with class labels. Top-down and bottom-up factors, with the former determined by current selection goals and the latter determined by physical salience, should be combined to direct attentional behavior. A recent review of attention models from a computational perspective can be found in Borji and Itti (2013), Borji, Sihite, and Itti (2012). Saliency models have been developed for eye fixation prediction and salient object detection. The former focuses on identifying a few fixation locations on natural images, which is important for understanding human attention. The latter, also called salient object segmentation, is used to accurately detect where the salient object should be, which is useful for many high-level vision tasks (Yang, Zhang, Lu, & Ruan, 2013; Borji et al., 2012).

Recently, the theoretical dichotomy of attentional control between top-down and bottom-up is challenged. The dichotomy fails to explain a growing number of cases in which neither bottom-up nor top-down can account for strong selection biases (Awh, Belopolsky, & Theeuwes, 2012). Thus, the top-down versus bottom-up dichotomy is an inadequate taxonomy of attentional control. Awh et al. (2012) proposed *selection history* (including two classes of 'history' effects, i.e., selection and reward history) as a third category of control by explicitly distinguishing current goals from selection history effects. A 'priority map' which they still highlighted integrates three distinct categories of selection bias: the observer's *current selection goals*, *selection history*, and *physical salience* of the items competing for attention. Acknowledging selection history as a third category of control can clarify many ongoing debates and can make clear large swaths of selection phenomena that are unrelated to current selection goals and physical salience. This concept model is a breakthrough to traditional prominent models of attentional control.

In this study, we propose a general computational framework for detecting specific salient objects (e.g. cars and pedestrians) in images beyond top-down and bottom-up mechanisms and verify qualitative and quantitative effects of current selection goals and selection history in our experiments. Salient objects are detected by directly measuring the saliency of an image window in the original image and the well established sliding window based object detection paradigm is adopted.

The main contributions of this study are our integrative computational framework and experimental conclusions. Our experimental results on challenging object detection datasets demonstrate that physical salience generates a bottom-up saliency map for highlighting the salient regions of an image. The main effect of the selection history is to concentrate on salient objects, the current goal has a strong effect in detecting correct salient objects. Experiments also indicate that there is competition among selection history, current goal and physical salience to detect correct salient objects.

The rest of this paper is organized as follows. Section 'Related works' introduces related works. Our computational framework is described in Section 'Our salient object detection framework'. Experimental results and comparisons are presented in Section 'Experimental results', and conclusions are given in Section 'Conclusions'.

Related works

In this study, our primary goal is to present a general computational framework for detecting salient objects in images. Selection history is integrated with current goals and physical salience to compose an integrative framework.

The objectness measure (Alexe, Deselaers, & Ferrari, 2010) quantifies how likely an image window contains an object of any class. Each outputting image window is endowed with an objectness score to measure how likely this window contains a salient object. It uses several existing image saliency cues (including a novel 'superpixels straddling' cue to capture the closed boundary characteristic of objects), and greatly reduces the number of windows from an image according to their objectness distribution. We use it to generate physical salience and quantify how likely an image window contains a salient object.

Feng, Wei, Tao, Zhang, and Sun (2011) proposed a salient object detection by composition. They presented a simple definition for window saliency, i.e., the cost of composing the window using the remaining parts of the image. Based on a segment-based representation, the window composition cost function can be evaluated by a greedy optimization algorithm.

LabelMe (Russell, Torralba, Murphy, & Freeman, 2008) is a web-based image annotation tool that is used to label the identity of objects and where they occur in images. We use the HOG (Histograms of Oriented Gradient) descriptor from the LabelMe toolbox and extend it to extract image features of current goal, selection history and sampled image windows.

Our salient object detection framework

Fig. 1 is a schematic diagram of our framework. Selection history, physical salience, and current goal are three distinct sources of selection biases to accomplish salient object detection.

(1) *Selection history*. This category of control is intended to represent 'history' effects which shape the overall landscape of the observer's selection biases.