Neurocomputing 167 (2015) 390–405

Contents lists available at ScienceDirect

Neurocomputing

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

Salient object detection via contrast information and object vision organization cues

Shengxiang Qi, Jin-Gang Yu, Jie Ma*, Yansheng Li, Jinwen Tian

School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China

ARTICLE INFO

Article history:
Received 26 August 2014
Received in revised form
13 April 2015
Accepted 13 April 2015
Communicated by Luming Zhang
Available online 29 April 2015

Keywords:
Computer vision
Saliency detection
Contrast
Object vision organization
Image segmentation

ABSTRACT

As a popular topic, saliency detection has attracted lots of research interest benefiting for its valuable applications in computer vision and image processing. In this paper, we propose to delineate saliency by considering both the contrast and object vision organization. It consists of two stages. In the first stage, the primary element, *contrast saliency*, is acquired by measuring color contrast and color distribution with background prior and center prior to address the uniqueness and compactness of salient regions. In the second stage, inspired by the Gestalt principles of grouping from the study of visual perception, we take into account the properties of closure, proximity and similarity for object vision organization, and then provide the *object vision saliency filtering* to emphasize homogeneous saliency across similar and object-like regions. As for the task, a map called *object coverage confidence* is presented to express the closure by characterizing the probability of complete object areas with refined profiles, which is constructed by fusing multiple information prediction maps, implying probable closure areas of objects in different layers of an image. Experimental results on five publicly available benchmarks demonstrate that our model outperforms the state-of-the-art methods.

1. Introduction

Saliency detection, with its capability to locate important regions covering valuable contents in an image, is seemed as one of the vital mechanisms in selective visual attention [1]. Relying on such a preattentive processing in visual perception, humans can rapidly focus their eyes on parts of interest being full of information they need from scenes [2]. Since understanding the mechanism distinctly is difficult [3], much effort has been devoted to research this human vision system (HVS) [4,5]. Encouragingly, sequential achievements in this field has proven its worth in solving many computer vision problems including image classification [6,7], recognition [8,9], retrieval [10], segmentation [11,12], and compression [13]. Based on this observation, saliency detection would be beneficial for a wide range of practical applications.

Since the feature integration theory (FIT) was proposed by Treisman and Gelade [14], it has been respected as one of the most significant principles in visual attention for decades. According to the concept of saliency map expressing human conspicuousness of scene locations proposed by Koch and Ullman [4], Itti et al. [5] construct the milestone model of visual attention, which is the first complete implementation incorporating FIT to detect saliency. Inspired by this succeed, a lot of research interest in this field has been attracted and a variety of models have been developed since then [15–19], dealing with the task mainly from two modes: the bottom-up manner [20] and the top-down manner [21]. Bottom-up attention is stimulus-driven and based on characteristics of scenes, while top-down attention is task-driven and mainly determined by cognitive phenomena [3]. Meanwhile, according to different purposes, existing models establish saliency towards two aspects, one is eye-fixation prediction [22] and another is salient object detection [23]. As the concept implies, the former is intended for predicting locations which easily attract human attention in a scene, whereas the later often focuses on identifying the whole salient objects or regions from an image, always preparing for the application like object segmentation [24] and recognition [25]. In this paper, we elaborate our work on salient object detection under a bottom-up manner since it is independent with specific tasks showing a more general pattern in preattentive processing [2], and so as to computer vision.

For bottom-up processing, early works measure saliency from various perspectives, including biological-inspired imitation [5], information-based theory [15,26,27], discriminant mechanism [18,28], frequency domain [17,29], graph-based model [16,30], Bayesian framework [31,32], and sparsity decomposition [33]. These approaches have achieved good results to some extent, however, when regarding to complex scenes, they may often fail. This is well understood that traditional saliency is preferably only measured by contrast to highlight feature rarity or irregularity.

* Corresponding author.
E-mail addresses: shengxiang.qi@gmail.com (S. Qi), majie@mail.hust.edu.cn (J. Ma).

http://dx.doi.org/10.1016/j.neucom.2015.04.055
0925-2312/

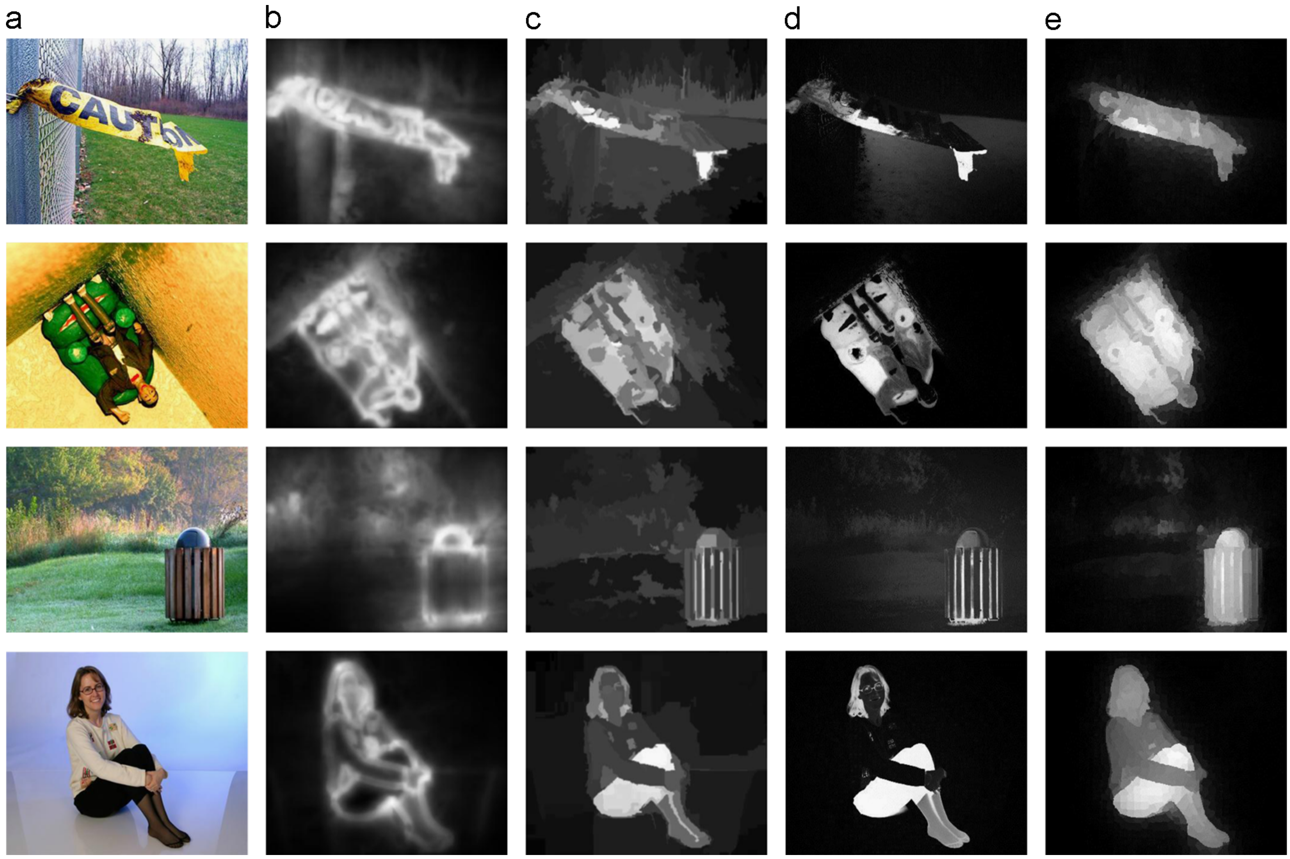

Fig. 1. Natural images (a) with their saliency maps produced by: (b) CA [46], (c) RC [47], (d) SF [48], and (e) Ours.

In fact, the interior of foreground object may contain various features (see Fig. 1(a)), leading to the invalidation of contrast-based measurement. To overcome this shortage, recently, object-based attention [34–36] is introduced into saliency modeling, which argues that saliency should be computed on regions formed by perceptual grouping procedures according to basic Gestalt principles [37,30]. Under this guidance, saliency detection could be improved by a more complete constraint with the construction of object organization [38–40].

In this paper, we elaborate on providing a new salient object detection model inspired by the Gestalt principles [37]. Although this concept is not novel and has been used before in measuring saliency [41,42], we are completely different. Formally, in our work, saliency is estimated by combining both contrast information and object vision organization cues to characterize the visual grouping of object. It emphasizes homogeneous saliency across similar and object-like regions via an optimization considering properties of contrast, closure, proximity and similarity. To this end, the proposed model consists of two stages: the first stage for expressing contrast and the second stage for describing object vision organization.

In the first stage, we first weight regions adjacent to image boundaries as the spatial background prior as the work in [43,44]. Then, we compute the color contrast by measuring color dissimilarity between all regions to backgrounds in feature space, to reflect the uniqueness of salient regions, and calculate color distribution through evaluating the dispersiveness of color locations in spatial domain, to address the compactness of salient regions. After this, a *contrast saliency* (CS) map is generated via integrating the color contrast and color distribution with center bias. CS makes up the primary element for saliency measurement. In the second stage, we first present a map called *object coverage confidence* (OCC) to express the closure by characterizing the probability of complete object areas with refined profiles. It is constructed by fusing multiple information prediction maps which imply probable closure areas of objects in different layers of an image. These information maps are computed by multi-layer segmentations attributing to a bipartite graph partitioning approach [45]. Due to the coverage of probably entire objects, OCC provides the visual character of closure suggesting that objects are habitually perceived as being whole by individuals [37]. Then, given the CS and OCC, we propose the *object vision saliency filtering* (OVSF) to assign homogeneous saliency values across similar and object-like regions. Specifically, three terms are contained in OVSF, namely a constraint on contrast (with CS), a constraint on closure (with OCC), and a smoothing term. The smoothing term is to force homogeneous saliency for close and similar regions, which is in accordance with the visual proximity and similarity in human perception [37]. Accordingly, OVSF takes into account effects of contrast, closure, proximity and similarity simultaneously. As a result, the final saliency is obtained by optimizing an expression of energy minimization with an iterative solution. Some saliency results of typical models and ours are shown in Fig. 1, as can be seen that compared to other maps with heterogeneous values, our saliency maps could highlight objects as being whole even for complex scenes. For overview, Fig. 2 shows the framework of our proposed model.

In summary, major contributions of the proposed model cover four aspects:

1. A novel measurement, *object coverage confidence* (OCC), is presented to express the visual character of closure, which characterizes the probability of complete object areas with refined profiles, to indicate the entire closure object as being whole.
2. An optimization, *object vision saliency filtering* (OVSF), is proposed to estimate saliency on object, which integrates both contrast and object vision organization cues including properties of contrast,