Neurocomputing 124 (2014) 57–62

Contents lists available at ScienceDirect

Neurocomputing

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

Letters

Background contrast based salient region detection

Huiyun Jing [a], Xin He [b], Qi Han [a,*], Xiamu Niu [a]

[a] Department of Computer Science and Technology, Harbin Institute of Technology, China
[b] National Computer Network Emergency Response Technical Team / Coordination Center of China, Beijing, China

ARTICLE INFO

Article history:
Received 16 December 2012
Received in revised form
1 April 2013
Accepted 11 July 2013
Communicated by Paolo Remagnino
Available online 24 August 2013

Keywords:
Salient region detection
Background contrast
Eye fixations

ABSTRACT

This paper proposes a novel salient region detection method based on background contrast. Salient regions are widely considered as those distinct to the image background. However, owning to the lack of information about the image background, existing methods measure saliency of an image pixel/region using its contrast to local neighborhoods or the entire image rather than the image background. Inappropriate contrast region leads to difficulty in highlighting the whole large salient region. In this paper, we discover that the absence of eye fixations provides an important clue to the image background. Regions without eye fixations are very likely the image background. To further acquire the spatial extent of image background, the complementary area of the convex-hull of eye fixations is used to represent the possible image background. Then we measure the saliency of each region by computing its contrast to the estimated image background. The experimental results demonstrate that our approach outperforms previous state-of-the-art methods.

1. Introduction

Humans can rapidly and accurately identify salient regions in their visual fields. Simulating such an ability in machine is critical to make machine handle visual content like humans. In the past decades, extensive research efforts have been conducted to construct the computational model for detecting salient region. Most of these methods [1–4] intend to predict human eye fixations, where the eye fixates for a while to gather and process interesting information. Itti et al. [1] directly summed multi-scale center-surround responses of visual features to generate saliency map. Murray et al. [2] integrated the information at multiple scales by performing inverse wavelet transform on the scale-weighted center-surround responses. Hou et al. [3] measured the saliency of each pixel in the frequency domain. Ban et al. [4] introduced psychological distance into the prediction of eye fixations. However, eye fixations are just some spatial discrete points, they cannot accurately represent the whole salient region [5,6].

Recent years have witnessed more interest in salient region detection, which is useful in applications such as image segmentation [7], object detection [8], and content aware image resizing [9]. We focus on data-driven salient region detection.

It is widely agreed that salient regions are distinct to the image background [5,6,10]. However, due to the lack of any knowledge about the image background, current methods measure saliency of an image pixel or region using its contrast with respect to local neighborhoods [11,5,6,12] or the entire image [13–17]. Though these methods have achieved promising results, they still have difficulty in highlighting the entire salient region uniformly (Fig. 1(1)). Besides, in the presence of large salient objects or complex backgrounds, they may fail to correctly highlight the salient regions (Fig. 1(2)).

Though the eye fixations cannot mark the whole salient region [5,6], the absence of eye fixations provides a clue to image background. Observing the outside world, we move our fovea in the retina to specific salient regions to fixate on them [18]. During the fixation, we get a detailed impression of them. We then move, or saccade, to the next location of interest. Thus human eye fixations always fix on the interesting parts of image [19], seldom on the image background. The regions without eye fixations are very likely the image background.

In this paper, guided by the eye fixations, we estimate the spatial extent of image background. Then we measure the saliency of each small image region using its contrast to the estimated image background. In Section 2 of the paper, the limitations of existing methods are described. Section 3 introduces the proposed salient region detection approach. Experimental results and conclusion are respectively presented in Sections 4 and 5.

2. Limitation of the existing methods

According to the results from perceptual research [10], contrast is the most influential factor in low-level visual saliency. Current methods detect saliency by calculating the contrast of each image region relative to its certain context. Based on the extent of the

* Corresponding author.
E-mail address: hanqi_xf@hit.edu.cn (Q. Han).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.07.043

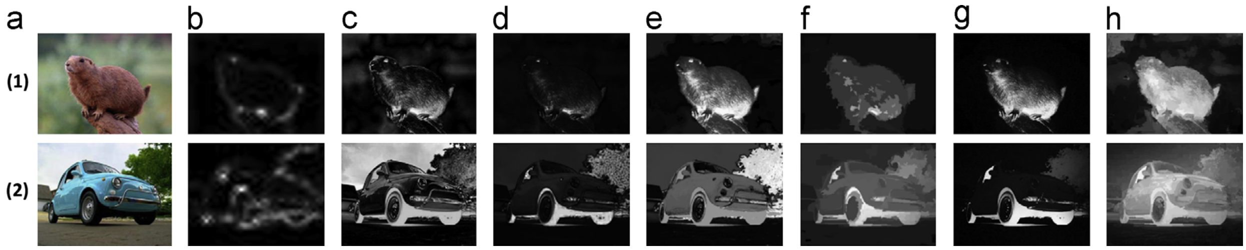

Fig. 1. Saliency maps computed by six state-of-the-art methods (b–g), and with our approach (h).

Fig. 2. Given input images (1), the corresponding eye fixations are represented as red dots (2). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

context, such methods can be classified as local methods [6,11,12] or global methods [13–16].

Local methods estimate the saliency of each image pixel/region using its contrast with respect to a small local neighborhood, e.g., computing the center-surround differences [11,12], or linearly combining multi-scale contrast in a Gaussian pyramid [6]. Because of carrying out local contrast, such methods intend to concentrate on the regions near edges, instead of highlighting the whole salient regions (Fig. 1(b)). It can be called "uniformly highlighting the salient regions" problem.

Global methods take the entire image as the context to compute the saliency of each pixel/region. Zhai and Shah [14] used the luminance contrast (LC) of each pixel to all other pixels. Achanta et al. [13] proposed a frequency-tuned (FT) approach that selects the appropriate spatial frequency range of image to detect salient region. Cheng et al. [15] proposed two global contrast-based salient region detection methods, including a histogram-based contrast (HC) method and a region-based contrast (RC) method. Perazzi et al. [16] presented Saliency Filters (SF) method, which formulates the global contrast measures as high-dimensional Gaussian filters. These global methods alleviate the problem of "uniformly highlighting salient regions". However, when the salient region is very large or the background is complex, such methods intend to highlight the background instead of the salient region (Fig. 1(2)), which is called "highlighting the background" problem.

Salient regions are distinct to the image background. However, existing methods use local neighborhood or the entire image as the contrast region to compute saliency. In essence, inappropriate contrast region of existing methods results in the above-mentioned problems. Nevertheless, lacking information about the spatial extent of image background makes it impossible for existing methods to use background as the contrast region. We find out that the absence of eye fixations provides an important clue to the image background. Thus we try to acquire the spatial extent of image background by detecting the regions without eye fixations. Once the image background is identified, the contrast to background can be computed to measure saliency.

3. The proposed approach

3.1. Overview of the proposed approach

Owing to the conflict between limited cognitive resources and large amounts of visual input data, only some important aspects of the input are selected for further detailed processing [20]. To accomplish this task, Human Visual Attention (HVA) is employed to align a salient region with the fovea [18], which is located at the center of the retina and has the highest resolution. It involves the detection of rapid eye movements called saccades and stationary phases called fixations. During the period of fixation, the eye-gaze holds steady on the fixation location to inspect the salient region. It allows detailed processing of the important parts of visual input.

Eye fixation locations are consistently specific to salient regions/objects [19]. Furthermore, they are mostly within the salient regions, and seldom on the background. In other words, the regions without eye fixations are most probably the image background. In Fig. 2(2), we note that the salient regions (the man sitting on sofa, castles, insect and bird, girl, the face of a man) contain the most of eye fixations and image background covers the greatest part of the regions without eye fixations.

Based on the above-mentioned discovery, the complementary area of the convex hull of all eye fixations can be regarded as the possible image background (Fig. 3(c)). The convex hull of a point-set is the smallest convex set that contains all of the input points [21]. The area where eye fixations locate could be compactly surrounded by the convex hull of eye fixations. Thus the complementary area of the convex-hull of eye fixations would closely represents the possible image background.

Once the image background is determined, the contrast to background can be computed. The original input image is firstly segmented into subregions. Then the saliency value of each subregion can be obtained by computing the weighted sum of its contrasts to all the subregions belonging to the image background. After executing the above processes, the saliency map indicating the salient region is finally obtained (Fig. 3(d)).