Signal Processing: *Image Communication* 29 (2014) 424–433

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Visual saliency detection based on region descriptors and prior knowledge

Weining Wang [a], Dong Cai [a,*], Xiangmin Xu [a], Alan Wee-Chung Liew [b]

[a] School of Electronic and Information Engineering, South China University of Technology, China
[b] School of Information and Communication Technology, Griffith University, Australia

ARTICLE INFO

Article history:
Received 12 July 2013
Received in revised form
7 December 2013
Accepted 16 January 2014
Available online 31 January 2014

Keywords:
Visual saliency
Visual attention
Object detection
Nonlinear feature integration

ABSTRACT

Visual saliency detection not only plays a significant role, but it is also a challenging task in computer vision. In this paper we propose a new method for saliency detection. It incorporates visual features and spatial information with a guidance of prior saliency knowledge. To provide more accurate visual cues, region descriptors are introduced for image segments by computing two saliency measures, namely feature distinctiveness and spatial distribution. In contrast to previous models which linearly combine basic features for visual cues, we provide nonlinear integration of features. In addition, by taking the advantage of the prior saliency distribution obtained from a convex hull of salient points, we heighten the contrast of fore- and background. Thereby we enhance the final saliency map that uniformly covers the salient objects, while tone down the nonsalient background. Experimental results on a benchmark dataset show that our saliency detection model performs favorably against the state-of-the-art approaches. A detailed experimental evaluation demonstrates that our algorithm excels at saliency detection in cluttered images.

1. Introduction

When human observes a given image, their fixations vary among different areas. Those areas attracting most interest and attention usually contain the most valuable information. They will be selected by a visual system for further processing. It is believed that visual saliency is closely related to those most informative areas, especially during free viewing in a bottom-up way. Therefore saliency detection is an important preprocessing step for computer vision tasks. Nowadays, the achievement of saliency detection has been applied in many fields such as image segmentation [1,2], object recognition [3], video retargeting [4], and aesthetics evaluation [5].

Saliency detection methods can be categorized into fixation prediction based and salient object oriented methods. Fixation prediction based methods are strongly related to biological models. Their processes simulate how human fixations locate on an image. Hence they tend to produce fovea-sized salient regions on the saliency map. On the other hand, object oriented methods generate maps such that salient regions uniformly cover the whole objects. One of the frequently adopted principles is computing the saliency of a segment with respect to the dissimilarity against its counterparts.

Recent object oriented research has shown their capability by utilizing low-level visual cues with basic visual features such as mean color or the region centroid. Even though these low-level cues have demonstrated favorable results, they are ill-suited in some cases e.g., the luminance variation on an object. To tackle this issue, some methods attempted to combine different features together. However, most of them simply integrated the saliency responses in a

* Corresponding author. Tel.: +86 13430281617.
E-mail address: c.dong04@mail.scut.edu.cn (D. Cai).

0923-5965/$ - see front matter © 2014 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.image.2014.01.004

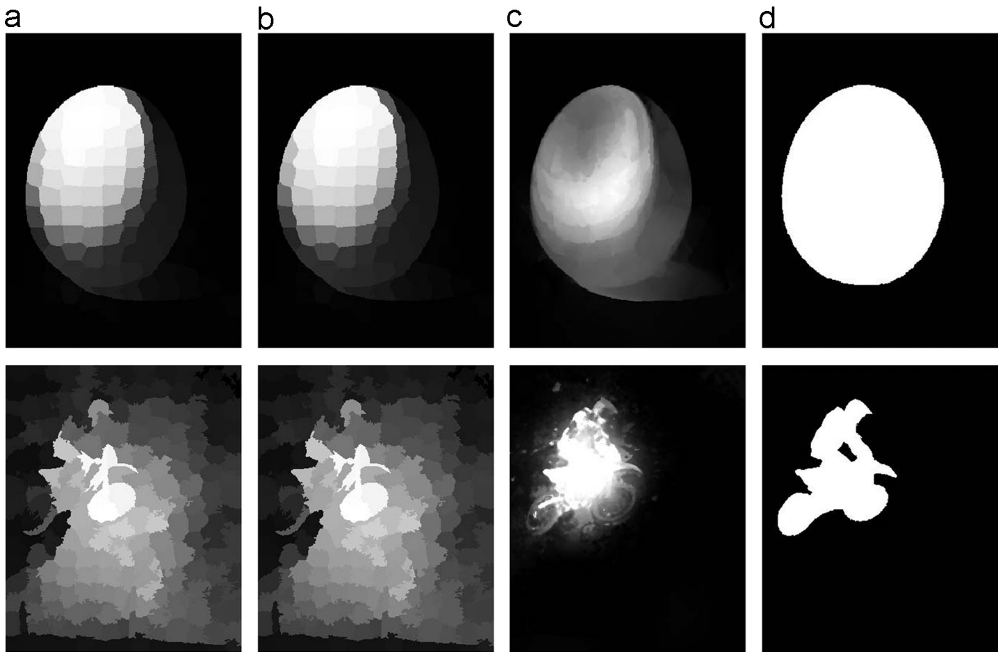

Fig. 1. Examples of object with luminance variation and object in complex environment. From left to right: (a) input images; (b) saliency maps of GBMR [14]; (c) our saliency maps; (d) the ground truth.

linear way [6,7]. Few studies evaluated the significance of different features to the overall saliency. Besides, since most previous region based methods are lack of prior knowledge about the salient region, it is difficult for them to discriminate the salient region from cluttered background. Some previous works [8–10] incorporated a center-bias prior map into their framework in order to improve their performances. However, because not all salient regions are located at the center, this approach is not applicable universally. Instead, it may weaken the performance of sophisticated models [11].

To address those problems remaining in existing object oriented methods, we introduce a region descriptor yielded by nonlinear feature integration to characterize local visual cues of image segments and measure their differences. Our feature integration is implemented by using covariance matrices of color and texture visual features. Additionally, we take the advantages of prior knowledge to improve our method. On one hand, we developed an adaptive approach of SLIC [12], a widely used pre-segmentation method, with prior knowledge of image complexity. This adaptive approach can reduce processing time because it determines the number of image segments according to image complexity. On the other hand, the contrast between the salient object and the background is heightened based on the prior saliency distribution obtained via a convex hull of salient points. This enhancement helps us gain an outstanding salient region and eliminate distracters in the background simultaneously.

In summary, the main contributions of this paper include the following: (1) We develop an adaptive approach of SLIC pre-segmentation method with the prior knowledge of image complexity. (2) We propose color- and texture-based region descriptors yielded by covariance matrices of multiple features to represent image segments. (3) We use the prior salient distribution based on a rough salient convex hull to enforce the contrast between the salient object and the background. Fig. 1 shows the comparison between our method and one of the widely used methods [13]. It illustrates that ours is robust to luminance variation and effectively highlights the salient object in the cluttered background.

The rest of this paper is organized as follows. Section 2 introduces the previous work in visual saliency detection field. Section 3 presents the details of the proposed saliency model. Experimental results and analyses are given in Section 4, and conclusion is shown in Section 5.

2. Related work

As mentioned above, the typical categories of saliency detection methods can be generally epitomized into fixation prediction and object oriented. Fixation prediction methods are inspired by biological principles. Itti et al. [15] proposed the seminal bottom-up saliency model derived from human visual selective attention mechanism [16]. It obtains saliency maps in different feature channels with a center-around operation and combines them linearly. Hou and Zhang [17] presented a saliency model counting for Fourier envelope and the differential spectral components, called the spectrum residual, to extract salient region. Erdem and Erdem [8] exert nonlinear integration of different features to describe image patches. Garcia-Diaz et al. [18,19] propose an adaptive approach achieved by decorrelation and contrast normalization. These fixation prediction based methods usually overemphasize local contrast and difference from neighborhood. They are more likely to produce spotlight saliency maps with low resolution and high saliency values on the object boundaries.