J. Vis. Commun. Image R. 25 (2014) 1823–1834

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Two-stage salient region detection by exploiting multiple priors

Qiang Fan, Chun Qi *

School of Electronic and Information Engineering, Xi'an Jiaotong University, Xian 710049, China

ARTICLE INFO

Article history:
Received 9 March 2014
Accepted 4 September 2014
Available online 16 September 2014

Keywords:
Superpixel isolation
Superpxiel distribution
Location prior
Multiple priors
Salient region
Two-stage
Image-feature-based
Precision/recall rate

ABSTRACT

Saliency detection has become a valuable tool for many image processing tasks, like image retargeting, object recognition, and adaptive compression. With the rapid development of the saliency detection methods, people have approved the hypothesis that "the appearance contrast between the salient object and the background is high", and build their saliency methods on some priors that explain this hypothesis. However, these methods are not satisfactory enough. We propose a two-stage salient region detection method. The input image is first segmented into superpixels. In the first stage, two measures which measure the isolation and distribution of each superpixel are proposed, we consider that both of these two measures are important for finding the salient regions, thus the image-feature-based saliency map is obtained by combining the two measures. Then, in the second stage, we incorporate into the image-feature-based saliency map a location prior map to emphasize the foci of attention. In this algorithm, six priors that explain what is the salient region are exploited. The proposed method is compared with the state-of-the-art saliency detection methods using one of the largest publicly available standard databases, the experimental result indicates that the proposed method has better performance. We also demonstrate how the saliency map of the proposed method can be used to create high quality of initial segmentation masks for subsequent image processing, like Grabcut based salient object segmentation.

1. Introduction

The human vision system can rapidly and accurately select some important regions to focus on when we are looking at something, we call such ability *Visual Saliency*. Saliency detection is an easy task for human beings while it is hard for computer. However, such ability is very important for many computer vision applications, like image retargeting [1], image classification [2], object recognition [3], summarization of a photo collection [4]. Therefore, saliency detection problem has been extensively studied in signal processing, machine learning, and biological literature (e.g. [5–8]). Early work by Koch and Ullman [9] and subsequent attention theories proposed by Itti et al. [8] and others suggest two categories of visual saliency: bottom-up, data driven, task independent visual saliency and top-down, goal driven, task dependent visual saliency. In this paper, we focus on the first category.

According to the applications, the saliency detection methods can mainly fall into two groups. In the early time, people pay more attention to detect the salient points which attract us at the first glance [8,7,5,10]. Recently, with the development of the application based on object-level saliency detection methods, like content-aware image resizing [1], object detection [11], more methods focus on detecting a salient object. In this paper, we aim at detecting the object level saliency.

As stated in [12], all the bottom-up saliency methods rely on the assumptions or priors on the properties of objects and backgrounds. The most fundamental prior is "appearance contrast between object and background is high", which is called *contrast prior*, and all the methods use this prior explicitly and implicitly. Contrast prior mainly explores the difference between the features of a pixel or a region and its local neighborhood or the whole image. This type of methods have achieved great success but still have some limitations, like "object attenuation" problem, which means that the boundaries of the salient object can be found well but the object interior is attenuated. In order to solve this problem, Wei et al. [12] solve the saliency detection problem from a different view, they study "what the background should look like" instead of "what the salient object should look like", and propose two priors about background. They assume the image boundary is mostly background, and image patches in the background can be easily connected to each other, namely *boundary prior* and *connectivity prior*. Based on the contrast prior and two background priors, the authors propose a Geodesic saliency measure, that is, the saliency value of an image patch (superpxiel) is the length of its shortest path to the virtual background node. They propose an

* Corresponding author.
E-mail address: qichun@mail.xjtu.edu.cn (C. Qi).

http://dx.doi.org/10.1016/j.jvcir.2014.09.003
1047-3203/

algorithm to realize the idea and extend this algorithm to superpixel.

However, using these priors only is not enough for a good detection of saliency. Take Fig. 1 for example, it shows saliency maps from three recent representative contrast based saliency detection methods [13,14,12]. From this figure we can see, these methods cannot always detect right salient regions. Take the first image for an example, saliency maps obtained by three methods cannot achieve low saliency values in the textured backgrounds, which cannot meet our will. This phenomenon encourages us to think about this problem again, apart from these priors, what else contributes to the saliency.

After studying the previous literatures, we have found that there are mainly three factors that influence the salient region detection problem, isolation, distribution, location prior. Isolation explores how different the feature (e.g. color) between a patch (region) or a pixel and the background. Distribution explores how the features (e.g. color) of a patch (region) or a pixel distribute in the image. Location prior explores where are the salient regions located at.

As for location prior, many previous works have been done. Judd et al. [15] analyze the human fixation points and find most of the human fixation points are near the center of the image, this bias is attributed to the setup of the experiment where users are placed centrally in front of the screen and also to the fact that human photographers tend to place the salient object in the center of photographs. They find that using a Gaussian blob centered in the middle of the image as the saliency map produces good result. Based on this finding, some methods like [16,17] use a central bias as a weight to find out the salient region. However, for some images in which the salient objects are not near the center of the image, this prior is not very reasonable. Goferman et al. [13] find a previous psychological evidence [18], it states that human beings will focus on one or several centers of the image. To simulate this, they firstly employ a global contrast based saliency detection scheme to get the saliency map, then the most attended pixels are selected if the saliency value of a pixel is above a threshold, the final saliency of a pixel is inversely proportional to the Euclidean distance between this pixel and the closest attended salient pixel. However, such selection of saliency pixel relies on the setting of parameters.

In this paper, we propose a salient region detection method which aims at overcoming the above problems. In the proposed algorithm, we use the Geodesic based saliency similar as [12] as the isolation measure, this method uses not only contrast prior, but also two background priors, and shows better performance compared with previous contrast based methods. However, using these priors only is not enough. Take the first image in Fig. 1 for example, the blue texture background has a high contrast with the white background, so the saliency value of the blue texture background computed by Geodesic based saliency is high (some regions are even higher than the book), this result cannot meet our will. We will regard the book as the most salient object, and the large spread texture background is less salient. To overcome this problem, we incorporate a distribution measure, that is to say, the patch or pixel whose pattern (e.g. color) spreads widely will receive a low saliency value. Some previous methods have used this measure (e.g. [19,20]). In this work, we use the element distribution measure similar with that proposed by Perazzi et al. [19] to realize this assumption.

As for location prior, both of the two methods we mentioned before are not very reasonable. We consider that the salient object will not always located at the center of the image, and the regions closed to the salient object will be focused more than faraway regions. When looking at an image, human beings will be attracted by the regions that stand out as salient regions, then they will focus more on these regions than faraway regions. In this paper, we introduce a new location prior measure to simulate this process. Particularly, we first compute the centroid of the image-feature-based saliency map which is computed before the location prior measure, then we regard this centroid as the foci of attention, and use a Gaussian falloff weight to emphasize the area close to the foci of attention.

Fig. 2 shows the overview of the proposed framework. First, the image is segmented into superpixels. In the first stage, two measures are used to measure the isolation and distribution of each superpixel, we consider that both of these two measures are important for finding the salient regions, thus the image-

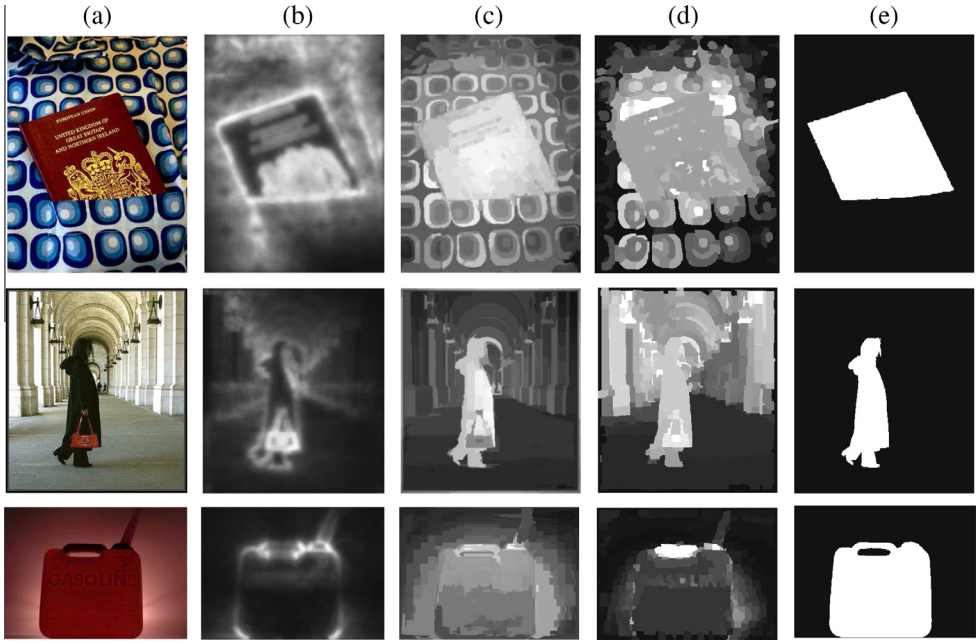

Fig. 1. Saliency maps of three previous representative contrast based methods on three example images. (a) Input images. (b)–(d) Results from method of [13,14,12]. (e) Ground truth.