Pattern Recognition 47 (2014) 1740–1750

Contents lists available at ScienceDirect

Pattern Recognition

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

Salient object detection via local saliency estimation and global homogeneity refinement

Hsin-Ho Yeh [a], Keng-Hao Liu [a,*], Chu-Song Chen [a,b,**]

[a] Institute of Information Science, Academia Sinica, Taipei, Taiwan
[b] Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 19 February 2013
Received in revised form
29 October 2013
Accepted 13 November 2013
Available online 23 November 2013

Keywords:
Salient object detection
Local contrast
Global homogeneity

ABSTRACT

This paper presents a new hybrid approach for detecting salient objects in an image. It consists of two processes: local saliency estimation and global-homogeneity refinement. We model the salient object detection problem as a region growing and competition process by propagating the influence of foreground and background seed-patches. First, the initial local saliency of each image patch is measured by fusing local contrasts with spatial priors, thereby the seed-patches of foreground and background are constructed. Later, the global-homogeneous information is utilized to refine the saliency results by evaluating the ratio of the foreground and background likelihoods propagated from the seed-patches. Despite the idea is simple, our method can effectively achieve consistent performance for detecting object saliency. The experimental results demonstrate that our proposed method can accomplish remarkable precision and recall rates with good computational efficiency.

1. Introduction

Identifying the visually attentive area in an image without a priori knowledge about the scene is the fundamental to our vision. Let us take a look at the image shown in Fig. 1(a) as an example. In this figure, people can easily identify the most salient words appeared. This image was generated by wordle.net, a web-tool converting a textfile into a figure. It extracts the keywords frequently occurred in the textfile and generates a figure consisting of those words for visualization, where a higher-frequency word is generated to occupy a larger area with more distinctive colors or brightness, which thus presents higher "visual saliency." Note that such a salient-words synthesis process can be seen as an inverse problem of salient object detection. Since the textfiles to be submitted are arbitrarily composed of web users, there is no prior knowledge about the frequency of the words (almost all words, except several stop words, have the same opportunities of becoming the keywords). Nevertheless, due to the bottom-up nature of saliency region detection, people can still find the salient objects even when there is no high-level or prior object-categorical information.

Salient regions, referred to as the image part attracting most human's attention, are fundamental for many high-level tasks in computer vision. Recent advance shows that salient regions provide useful information for image segmentation [1], object recognition [2,3], and motion detection [4]. They also contribute to improve the performance of many image quality metrics [5] and aesthetics value assessment [6–8]. Early studies of computational saliency [9,10] used biology-inspired models to simulate the selection mechanism of Human Vision System (HVS). In such model, visual input is decomposed into a set of feature maps by several pre-attentive filters based on image features such as colors or intensity. The saliency in a feature map is regarded as the competition result of neurons (or locations) in the spatial domain, where the prominent neurons significantly differ from their surrounds can survive. Then, the saliency intensities in all feature maps are integrated into a final map.

Later, several studies were introduced to enhance the saliency detection performance of biology-inspired models. Gopalakrishnan et al. [11] and Wang et al. [12] employed random walk to model the competition process. Valenti et al. [13] designed several new pre-attentive filters based on edge and color responses. Wang et al. [14] learned the dictionary from a large-scale set of images for estimating the intrinsic and the extrinsic saliency. Murray et al. [15] explored the integration step of the feature maps for further refinement purpose.

In practice, the performance of most biology-inspired methods is still limited, which could be owing to the difficulties encountered in choosing or learning the filters. Thus, various approaches

* Corresponding author.
** Principal corresponding author at: Institute of Information Science, Academia Sinica, Taipei, Taiwan. Tel.: +886 2 27883799x1310.
E-mail addresses: hhyeh@iis.sinica.edu.tw (H.-H. Yeh), keng3@iis.sinica.edu.tw (K.-H. Liu), song@iis.sinica.edu.tw (C.-S. Chen).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.11.015

Fig. 1. What are the salient keywords? (a) wordle.net provides us an efficient tool to summarize frequently appeared topic words in a document. Such a saliency-synthesis procedure poses an inverse problem of saliency detection. (b) The saliency detection result (per-pixel saliency degree) obtained by our proposed method. (c) The objective ground truth.

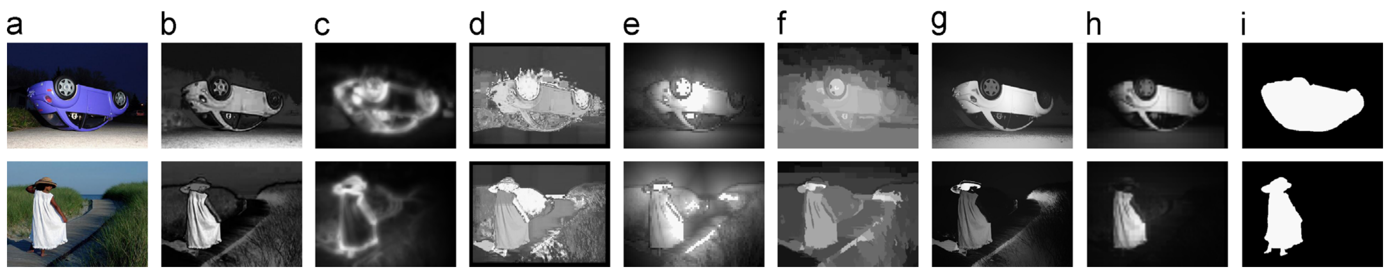

Fig. 2. Examples of salient object detection. (a) shows the original images. (b)–(f) demonstrate the state-of-the-art saliency maps from FT [16], CA [20], SS [18], CAS [21], RC [22], and SF [23]. The map obtained by our method LGA(h) is plotted as well. The (i) shows the ground truths (GT). From these results, the propose salient methods not only detect the salient regions well, but also reduce the noises better in the background clutter.

adopting other kinds of models or frameworks were proposed to resolve this dilemma. For instance, Achanta et al. [16] designed a band-pass filter to preserve a reasonable range of spatial frequency for detecting the salient regions. Lu et al. [17] found that the shape information, such as convexity and concavity, can reveal significant clues for locating the salient regions. In general, these approaches employ global information via frequency or shape analysis. However, they would often fail to detect small objects due to the fact that distinctive smaller regions are easily overlooked.

In order to solve the certain issues suffered from the global-based methods, another technique, referred to as local contrast, was proposed. It was derived from the concept of the so-called "center-surround" concept. Center-surround contrast has already been employed by Itti et al. [9,10], which reflects the local spatial-discontinuities of visual contrast. Local contrast, inspired from this mechanism, was thus developed to emphasize the uniqueness of a certain region by accentuating the contrasts to neighbors. It has been widely exploited by window-based approaches in [18,19]. Rahtu et al. [18] conducted an effective Bayesian formulation to measure the salient regions from color. To present the regionality, they separate a sliding window into the inner and collar sub-windows. Then a pixel's saliency is directly determined by the color contrast between the two sub-windows. This approach achieves nice results owing to the appropriate consideration of local contrasts. Klein et al. [19] further extended it by using Kullback–Leibler divergence. However, without knowing the object size, the methods of local contrast have to change their window size to locate the saliency for different scales of object. The problem caused by object-sized variation would degrade their performance significantly in many cases.

Based on the above-mentioned reasons, how to relax the spatial constraints of window-based approaches could be a main issue of improving the performance. Goferman et al. [20] presented the spatially weighted color contrast (or called the surrounding contrast) for saliency detection. The surrounding contrast of a pixel is measured from the weighted distances (the color distances multiplied by the inverse of spatial distances) to the other patches. The saliency values of different scales are then averaged to obtain a single saliency map. Although their results are useful for certain applications such as image re-targeting, this approach usually suffers from the problem that the detection results are sensitive to the edges in an image, and tends to overlook homogeneous foreground regions.

From the above point of view, an ideal contrast-driven saliency detector should take both local perspective and global-homogeneous property into consideration. Many recent studies focused on the local saliency estimation based on the principles such as Rareness [18,24], Contrast [10,23], and Center-Surrounding [19,25]. They might neglect that a salient region usually differs from its neighborhood outside the region besides containing homogeneous parts inside the region. How to link locally distinctive patches/regions that could be homogeneously associated to the foreground is a crucial key.

In this paper, we propose to detect the salient objects from an initial local saliency estimation process, where several seed regions can be selected from the initialization. Then a homogeneity-growing competition process is designed to precisely locate the salient foreground region. Our approach can perform better than state-of-the-art results (such as [16,18,20–23], and is efficient to implement. Fig. 2 shows the examples of salient object detection results obtained by various methods.

This paper is organized as follows. Some of the related works are discussed in Section 2. The problem formulation and the proposed method are introduced in Sections 3 and 4, respectively. Experimental results and comparisons are shown in Section 5. Finally, conclusions and future works are drawn in Section 6.

2. More of the related works

In addition to the above-mentioned works, new methods about saliency detection is proceeding to be explored. For instance, Duan et al. [26] further addressed this problem by employing the PCA