Neurocomputing 203 (2016) 34–40

Contents lists available at ScienceDirect

Neurocomputing

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

Saliency detection by selective color features

Yanbang Zhang [a,b,*], Fen Zhang [a,c], Lei Guo [b]

[a] College of Mathematics and Information Science, Xianyang Normal University, Xianyang Shaanxi 712000, China
[b] School of Automation, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China
[c] Department of Automatic Control, Xidian University, Xi'an, Shaanxi 710071, China

ARTICLE INFO

Article history:
Received 9 January 2015
Received in revised form
14 March 2016
Accepted 27 April 2016
Communicated by L. Shao
Available online 6 May 2016

Keywords:
Visual attention
Saliency model
Region detection
Contrast-based saliency
Distribution-based saliency

ABSTRACT

This paper is concerned with selective color feature for detecting salient regions. In contrast to most existing methods related to detection in one color space, the proposed algorithm pre-segments the input image into superpixels in both RGBY color space and Lab color space. Next, to calculate color contrast we not only consider the local feature, but also compute the difference between the pixel and the whole image. In the meanwhile, based on the center-surrounding scheme, a new computational model of color distribution features is presented to detect salient regions. Finally, 2D entropy is employed as an evaluation criterion to select and integrate the proper color features. Experimental results show that the proposed method outperforms the state-of-the-art methods on salient region detection.

1. Introduction

Human visual system is an intelligent processing system, which can select the important parts referred to as salient objects in a complex visual environment while ignoring others. How to detect the salient objects quickly and precisely still is a challenging and key technique. Therefore, visual attention mechanism has received increasing interest in recent years, partly due to various applications, such as image retrieval [1], image segmentation [2], object tracking [3], adaptive coding [4], and so on.

So far, a variety of computational models for visual saliency detection have been proposed in physiology, cognitive science, computer vision, and other fields. From the viewpoint of information processing, these algorithms can be categorized into three classes [5], namely, bottom-up model, top-down model and hybrid model. Bottom-up model simulates human instinctive visual attention mechanism, which is rapid, preattentive, and stimulus driven, and always chooses the low-level features, for example color [6], intensity [7], texture [8] to extract salient objects that are different from their surroundings. A pioneering work, called Itti model [9], put forward a saliency detection model to select the most conspicuous locations as attended regions by combining three multi-resolution local features such as luminance contrast, chrominance contrast, and orientation contrast. Motivated by this work, Judd et al. [10] introduced several new features to characterize image content which included subband pyramids based features, 3D color histogram, horizon line detector, etc.

On the other hand, since top-down model is a goal-oriented, attentive, and task-dependent manner, which uses prior knowledge to detect what we need [11,12], it usually obtains more precise results than those of the bottom-up model at the sacrifice of computation. One of the most famous examples of top-down attention guidance came from Yarbus [13] in 1967, who showed that eye movements depended on the current task. Hybrid model combines the advantages of bottom-up model with top-down model, which attracts much attention recently, and many detection methods have been proposed [14] to obtain higher precision. In [14], a direct mapping was learned from bottom-up and top-down features to eye fixations by resorting to SVM, AdaBoost, k-Nearest Neighbor algorithm (KNN), and so on.

It is worth noting that it cannot predict salient objects before testing due to diversity and uncertainty of objects, so the existing research methods mainly focus on the bottom-up model. In this paper, we propose a bottom-up model to detect objects without resorting to content or specific prior knowledge about the targets. Solving the salient object detecting problem can benefit many computer vision tasks and application.

* Corresponding author at: School of Automation, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China. Tel.: +86 029 33720773.
E-mail address: zhyb@mail.nwpu.edu.cn (Y. Zhang).

http://dx.doi.org/10.1016/j.neucom.2016.04.005
0925-2312/

2. Related work

When we look at an image, some objects attract our attention due to the fact that their colors are different to their surrounding areas. Since human vision is very sensitive to color, color plays an important role in detecting saliency. Achanta et al. [15] defined the color difference between every pixel and the mean value of whole image as saliency value. Achanta and Susstrunk [16] extended the work [15] to deal with each pixel in a symmetric surround rather than the whole image in the presence of local features. In the spectral domain, Hou and Zhang [17] proposed a saliency detection method by using the spectral residual of the log spectrum of an image. The model in [18] detected image saliency by analyzing the contrast in the spectral and spatial domain simultaneously, but did not involve the feature of the color distribution. Whether the pixel belongs to the salient object depends not only on pixel value of color, but also on the distribution and position of the pixel. Hence, some models that combine color contrast with color distribution to detect salient region have received more attention [19–21]. Taking advantage of color contrast and color distribution, a salient object detection model was offered in [19]. Perazzi et al. [20] proposed a contrast-based saliency model, which was consistently formulated as high-dimensional Gaussian filters. Unfortunately, they only considered the local contrast feature in Lab color space, while ignoring the information about the global ones. Based on spatial-color constraint and similarity distribution constraint, Xu et al. [21] showed a method that measured pixel-level saliency and region-based saliency.

It is well known that feature combination is a crucial factor to improve the performance of saliency detection. In 2010, Atrey et al. [31] reviewed many state-of-the-art fusion strategies, for example, addition, multiplication, mean value, and so on. As pointed out by Zhang et al. [18], 2D entropy is a very useful tool to select the features that are combined to obtain the final saliency.

Inspired by Zhang et al. [18] and Perazzi et al. [20], we study the features of the color contrast and the color distribution in both RGBY and Lab color spaces, and propose a bottom-up model to detect the salient regions in this paper. Unlike the aforementioned methods in [20], the contrast features of local and global are considered simultaneously to detect the salient features. Moreover, we divide the combination process into two steps. First, we choose the proper channels to express the contrast and distribution features, and fuse them into the contrast feature map and the distribution feature map, respectively. Next, the final map is obtained by integrating the above results. Finally, comparison results show that the proposed method outperforms some existing ones.

3. The proposed method

In this section, we give a detailed description of the proposed saliency detection model. The section is organized as follows. The pre-segmentation model is introduced in Section 3.1. The color contrast feature and the color distribution feature are considered, respectively, in Sections 3.2 and 3.3. Feature combination is presented in Section 3.4.

3.1. Pre-segmentation

Pixels that belong to the same region usually have the very similar color components, and then the computational complexity can be reduced by computing region based on contrast. Superpixel algorithms group pixels into perceptually meaningful atomic regions which have the advantages of replacement of the pixel, redundancy reduction, decreasing in complexity. In general, a good superpixel algorithm can improve the performance of the detection. Recently, Achanta [16] has proposed a new superpixel algorithm based on simple linear iterative clustering (SLIC), which adopts k-means clustering to generate superpixels, and performs an empirical comparison of the existing superpixel methods. Since human visual system is more sensible to opponent-color and Lab color space is close to the human visual attention, here, we use SLIC superpixels in RGBY color space and Lab color space, where RGBY is double opponency color space defined in [9]. In order to adapt to different salient objects, we adopt a multi-scale superpixel segmentation algorithm. For a given image, we segment it into 200, 300, 400, and 500 superpixels, respectively, and use

$$C_{i,n} = \frac{\sum_{I^C_{m,n} \in R_i} I^C_{m,n}}{\| R_i \|} \tag{1}$$

and

$$P_i = \frac{\sum_{I^P_{m,n} \in R_i} I^P_{m,n}}{\| R_i \|} \quad n \in \{RB, BY, I, L, a, b\} \tag{2}$$

to denote the average color and position of superpixels in the channel n, where $I^C_{m,n}$ is the value of the pixel at the point m in the channel n. $\| R_i \|$ is the number of pixels in superpixel region R_i. Note that, here, pre-segmenting the image into superpixels reduces redundancy and noise.

3.2. Contrast-based saliency

If an object's color is different from its surrounding areas, then it is more attractive, therefore, color contrast has a significant effect on saliency detection. Commonly used color features include saturation, intensity, hue, opponents, warm and cold colors, and so on. The measurement in human visual cortex [23] has shown that the strongest response is from red-green and blue-yellow stimuli, which therefore absorb more human attention. In addition, Lab color space is intended to perceptual uniformity. In this space, L component closely matches human perception of lightness, and a and b channels approximate the human chromatic opponent system. In this paper, motivated by the fact that one color system may not always work well [7,18], we adopt both RGBY space and Lab space, in which there exist three cases if a region is salient.

(a) The color of the region is different from that of its surrounding regions and the average color of the whole image as well.
(b) The color of the region is similar to the average color of the whole image, but different from that of their surrounding regions.
(c) The color of the region is similar with that of its surrounding regions, but different from the average color of the whole image.

Only when the color of the region is similar with those of both their surrounding regions and the whole image, this region is the background. Fig. 1 shows some examples of three cases.

Based on the above discussion, the saliency of a superpixel in the channel n is defined as

$$S_{SC}(i,n) = \exp\left(-\frac{|P_i - P^c|^2}{\sigma_1^2}\right) \sum_{\forall j \in I} W^{1,n}_{i,j} \left(D(C_{i,n}, C_{j,n}) + D(C_{i,n}, C_{mean,n})\right) \tag{3}$$

$n \in \{RB, BY, I, L, a, b\}$

with

$$W^{1,n}_{i,j} = \frac{1}{k} \exp\left(-\frac{|P_i - P_j|^2}{\sigma_2^2}\right) \exp\left(-\frac{|P_j - P^c|^2}{\sigma_3^2}\right) \tag{4}$$