J. Vis. Commun. Image R. 48 (2017) 238–253

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

# 3D saliency detection based on background detection

Hongyun Lin [a], Chunyu Lin [a,*], Yao Zhao [a], Anhong Wang [b]

[a] Institute of Information Science, Beijing Jiaotong University, Beijing Key Laboratory of Advanced Information Science and Network, Beijing 100044, China
[b] School of Electronic Information Engineering, Taiyuan University of Science and Technology, Taiyuan 030024, China

ARTICLE INFO

*Article history:*
Received 14 March 2016
Revised 20 January 2017
Accepted 25 June 2017
Available online 30 June 2017

*Keywords:*
3D saliency detection
Background detection
Envisioned background
Polynomial fitting
Salient region

ABSTRACT

Unlike 2D saliency detection, 3D saliency detection can consider the effects of depth and binocular parallax. In this paper, we propose a 3D saliency detection approach based on background detection via depth information. With the aid of the synergism between a color image and the corresponding depth map, our approach can detect the distant background and surfaces with gradual changes in depth. We then use the detected background to predict the potential characteristics of the background regions that are occluded by foreground objects through polynomial fitting; this step imitates the human imagination/envisioning process. Finally, a saliency map is obtained based on the contrast between the foreground objects and the potential background. We compare our approach with 14 state-of-the-art saliency detection methods on three publicly available databases. The proposed model demonstrates good performance and succeeds in detecting and removing backgrounds and surfaces of gradually varying depth on all tested databases.



## 1. Introduction

Saliency detection, or the attempt to detect visually salient regions in an image, has been receiving increasing interest because of its broad applicability in visual analysis. Two mechanisms of visual attention are typically distinguished: bottom-up and top-down [1]. Bottom-up attention is fast, subconscious, data-driven and independent of any specific viewing task, whereas top-down attention is slow, voluntary and dependent on both the viewing task and semantic information [2]. These two kinds of mechanisms interact with each other to affect human visual behavior [3–5]. Most previous works have focused on the bottom-up visual process because of the complexity of the top-down process.

The existing 2D saliency detection algorithms give priority to global and local contrast; Itti's model [6] and Cheng's RC [7] are typical examples. Achanta et al. [8] create a full-resolution saliency map through the superposition of multiple-scale regional contrasts. Goferman et al. [9] calculate saliency values under the premise that novel features and high-contrast regions are more salient and tend to be concentrated rather than dispersed in an image. The spectral residual approach [10] is based on the global contrast and builds a saliency map by analyzing the characteristics of salient regions in the frequency domain.

Several computational models of 3D visual attention have been investigated in the previous literature, and they can be divided into three categories: depth-weighting (DW) models, depth-saliency models and stereo-vision models [2]. DW models treat the depth as a weight by which to multiply a 2D saliency map [11]. Depth-saliency models treat depth saliency as additional information [12,13]. Depth features such as relative depth, the pop-out effect and depth gradients are first extracted from the depth map to create additional feature maps, which are then used to generate depth saliency maps (DSMs). Finally, these DSMs are combined with 2D saliency maps to obtain the final results. Unlike the other two approaches, which use depth maps directly, stereo-vision models take into account the mechanisms of stereoscopic perception in the Human Visual System (HVS) [14].

Most visual saliency models focus on summarizing the properties of salient objects. However, general properties shared by various categories of objects do not exist or at least are difficult to conclusively identify. In this paper, we convert this problem into the exploration of the properties of background regions. Our approach arises from several basic observations: Background regions are generally distributed along the boundaries of an image, often called the "boundary prior" [15–18]. By contrast, salient pixels are generally concentrated in the center of the image (center bias). Moreover, background regions can generally be divided into two categories (as shown in Fig. 7): distant background regions, which are far from the observer, and "depth-gradual surfaces", on which the change in the depth value is gradual. Based on the

* Corresponding author.
*E-mail address:* cylin@bjtu.edu.cn (C. Lin).

http://dx.doi.org/10.1016/j.jvcir.2017.06.011
1047-3203/

above properties of background regions, we are able to detect backgrounds using depth information and color cue. The detected background can be regarded as the baseline of an image, and the areas that stand out from the background are generally salient regions. Although the existing 3D saliency detection approaches utilize various features of depth maps to improve 2D saliency detection, we believe that the depth should be regarded not as supplementary information but rather as a dominant feature; in our approach, depth values serve as the key to distinguishing between backgrounds and foregrounds.

Different from previous works, we make full use of the spatial characteristic of depth to detect the background regions in an image. We then use the detected background to predict the characteristics of the occluded background regions rather than simply filling these occluded regions with the detected background characteristics. Our approach imitates the imagination/envisioning process of human beings to create a saliency map using our basic concept, as illustrated in Fig. 1. To obtain the saliency map, we rebuild the potential characteristics of the background regions that are occluded by foreground objects through polynomial fitting. As seen in Fig. 1, the baby and sofa are foreground objects, and the regions occluded by them are the background regions to be predicted. Although the occluded background is invisible, we can imagine/envision it based on the known background by applying polynomial fitting to obtain a curve from several known points; here, this process is referred to as the envisioning process. Thus, the obtained saliency map is equal to the original image minus the background image (containing both the detected and predicted backgrounds). In summary, our main contributions in this paper are as follows:

- We introduce 3D saliency detection approach that computes saliency value by detecting background instead of directly detecting salient region. Furthermore, we introduce depth and spatial distribution into background detection.
- We conclude five characteristics of the background: *boundary prior*, *boundary connectivity*, *homogeneity*, *distance* and *depth gradualilty*. These characteristics are developed to detect background regions (or pixels).
- We use polynomial fitting to imitate the human imagination/envisioning process (see Fig. 1) to rebuild the occluded background regions.
- A series of comparisons of our approach with 14 state-of-the-art saliency detection methods on three publicly available databases is presented to verify the effectiveness of the proposed scheme. We also analyze the impact of center bias on two databases with different ground truths and obtain useful guidance for the construction of improved saliency models.

The paper is organized as follows. Related work is presented in Section 2. Section 3 introduces our saliency detection method. Qualitative and quantitative experimental results are discussed in Section 4. Finally, the conclusions are presented in Section 5.

## 2. Related work

Several recent approaches have considered the exploitation of background in saliency detection. Xu et al. [19] use spatial similarity distribution of the background to detect saliency. They think the foreground objects and the widely spreading background can be "easily" distinguished because the foreground objects generally distribute in a focused way while the background distributes dispersedly. Wei et al. [20] exploit two common priors about backgrounds in natural images, namely*boundary* and *connectivity priors*, to provide more clues for saliency. Zhu et al. [21] further propose a robust background measure, called *boundary connectivity*. Wang et al. [22] calculate saliency via background and foreground seed selection. Chen et al. [23] think salient objects can be easily obtained only by measuring the differences between the original images and their corresponding background maps. The above mentioned literatures mostly exploit background information in 2D image. Our approach further extends to 3D scene and detects background regions by using both color cue and depth cue. It is hard to say for certain what is the background in the academic field, but we give its special definition for saliency detection, that is, the background is the widely spreading region that cannot catch human visual attention relative to the salient objects. On the basis of above related work and our observation on image, we conclude five characteristics of the background: *boundary prior* [15–18,20,21,24], *boundary connectivity* [21], *homogeneity* [20,23,24], *distance* and *depth gradualilty*.

## 3. Proposed scheme

Most previous methods have focused on the detection of salient regions; however, it is impossible or at least very difficult to extract general features of salient regions because of the diversity of target objects. By contrast, we can observe several general properties of background regions that can allow us to detect such regions based on the depth and spatial relationships between regions. After background detection, our approach employs the detected background to predict the potential characteristics of the background regions that are occluded by foreground objects. Finally, a saliency map is obtained based on the contrast between the foregrounds and the predicted backgrounds. A workflow diagram of the proposed model is presented in Fig. 2.

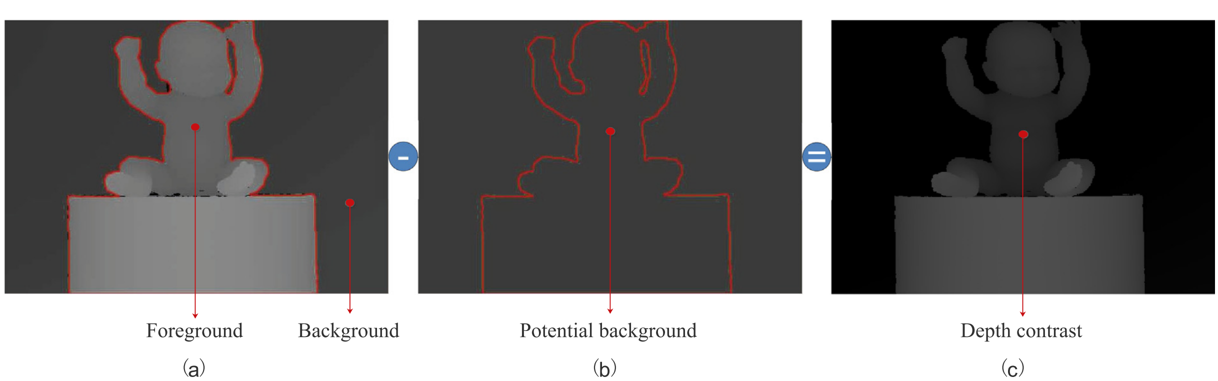


**Fig. 1.** The envisioning process and our basic concept (a – b = c) regarding the computation of DSMs. (a) Original depth image. (b) Background image (the detected and predicted background). (c) DSM. The envisioning process is the use of the detected background to predict the potential characteristics of the background regions that are occluded by foreground objects through polynomial fitting (see Fig. 11 for details).