Signal Processing: *Image Communication* 38 (2015) 3–14

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Image saliency estimation via random walk guided by informativeness and latent signal correlations

Jia Li [a,b], Shu Fang [c,1], Yonghong Tian [c,*], Tiejun Huang [c], Xiaowu Chen [a]

[a] *State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing 100191, China*
[b] *International Research Institute for Multidisciplinary Science, Beihang University, China*
[c] *National Engineering Laboratory for Video Technology, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China*

ARTICLE INFO

Available online 29 May 2015

Keywords:
Image saliency
Random walk
Informativeness
Latent correlation

ABSTRACT

Visual saliency is an effective tool for perceptual image processing. In the past decades, many saliency models have been proposed by primarily considering visual cues such as local contrast and global rarity. However, such *explicit* cues derived only from input stimuli are often insufficient to separate targets from distractors, leading to noisy saliency maps. In fact, the *latent* cues, especially the latent signal correlations that link visually distinct stimuli (*e.g.*, various parts of a salient target), may also play an important role in saliency estimation. In this paper, we propose a graph-based approach for image saliency estimation by incorporating both explicit visual cues and latent signal correlations. In our approach, the latent correlations between various image patches are first derived according to the statistical prior obtained from 10 million reference images. After that, the informativeness of image patches and their latent correlations are jointly considered to construct a directed graph, on which a random walking process is performed to generate saliency maps that pop-out only the most salient locations. Experimental results show that our approach achieves impressive performances on three public image benchmarks.

1. Introduction

Perceptual image processing, which aims to analyze images as human being does, is now becoming a hot research topic in the field of computer vision. In perceptual analysis, a key step is to locate important image content that demonstrates strong ability in capturing human visual attention. Toward this end, visual saliency can be estimated to quantize the importance of various image contents. By processing *visually salient* content with high priority, images can be efficiently analyzed, and the analysis results can better meet human perception.

In the past decades, hundreds of approaches have been proposed for visual saliency estimation. Among these approaches, most of them computed visual saliency as a kind of visual *rarity*, which were often measured by using explicit visual cues such as center-surround contrast and regional dissimilarity. For example, Itti et al. [1] proposed to estimate visual saliency by fusing multi-scale center-surround contrasts from multiple features. Harel et al. [2] represented images as graphs and detected salient pixels by defining the weights of graph edges as pixel dissimilarities. Moreover, some approaches tried to learn an optimal mapping mechanism from explicit visual cues to real-valued saliency scores. For instance, Navapakkam and Itti [3] proposed an approach to optimally combine local contrasts by

* Corresponding author.
E-mail address: yhtian@pku.edu.cn (Y. Tian).
[1] First two authors contributed equally.

http://dx.doi.org/10.1016/j.image.2015.05.005
0923-5965/

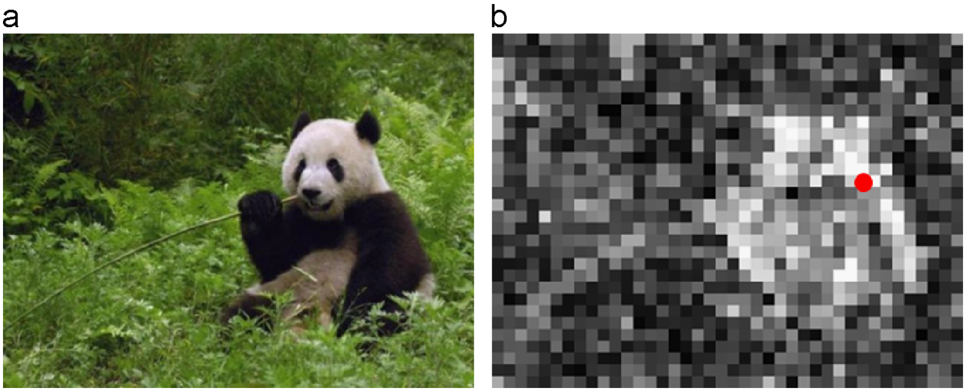

Fig. 1. Visual stimuli with distinct visual appearances may be inherently correlated. (a) A panda that consists of visually distinct parts; (b) the latent correlations between the 8×8 patch marked with red and all other patches.

maximizing the signal–noise-ratio. Zhao and Koch [4] proposed a boosting approach to train visual saliency model by fusing various visual features and their local contrasts. By focusing on explicit visual cues such as contrast and dissimilarity, all these approaches can pop-out targets and suppress distractors to some extent. However, such explicit cues are often insufficient to perfectly separate targets from distractors that share certain visual attributes. As a result, high saliency values may be wrongly assigned to distractors, resulting in "noisy" saliency maps.

To further separate targets from distractors, a feasible solution is to refer to additional saliency cues beyond the input image. By observing massive images, we find that various image contents can be inherently correlated, and such *latent signal correlations* actually link visually distinct stimuli (*e.g.*, various parts of a salient target, see Fig. 1 for an example). In visual saliency estimation, such latent signal correlations can work as a kind of prior knowledge that helps to further distinguish targets from distractors. Therefore, it is necessary to take such latent signal correlations into account in image saliency estimation.

Inspired by this idea, we propose an approach to estimate image saliency via random walk guided by informativeness and latent signal correlations. In our approach, the latent correlations between various image patches are first mined according to the statistical prior learned from 10 million reference images. These latent correlations between image patches, together with the patch informativeness, are then jointly considered to build a fully connected graph with directed edges and asymmetric weights. Under the guidance of informativeness and latent signal correlations, we divide image patches into three categories and adopt different random walking strategies between different types of patches. In this manner, the estimated saliency maps pop-out only the most salient locations while distractors can be well suppressed. Experimental results show that our approach achieves impressive performances in the comparisons with 13 approaches on three public image benchmarks.

Our main contributions are summarized as follows:

1. We incorporate latent signal correlations to facilitate the separation of targets and distractors. By exploiting the statistical prior obtained from 10 million reference images, the latent relationship between image patches can be well characterized so as to pop-out targets and suppress distractors.
2. We propose a graph-based approach that estimates image saliency via random walk guided by informativeness and latent signal correlations. In this approach, we divide image patches into three categories and apply different random walking strategies between different types of patches. In this manner, estimated saliency maps can pop-out only the most salient image locations while distractors can be well suppressed.

The rest of this paper is organized as follows: Section 2 briefly reviews related work, and Section 3 presents the learning process of latent signal correlations. In Section 4, we describe the details of the proposed saliency model. Experimental results are shown in Section 5, and the paper is concluded in Section 6.

2. Related work

In the past decades, many approaches have been proposed to estimate saliency in images and videos. In this review, we mainly focus on image saliency estimation. According to the visual cues used to estimate saliency, existing approaches can be roughly grouped into two categories, including the bottom-up category and the top-down category. We will briefly review approaches in these two categories from the perspective of visual cues they ever used.

2.1. The bottom-up approaches

The approaches in the bottom-up category mainly focus on the explicit visual cues that can be directly extracted from the input visual stimuli (*e.g.*, local contrast, dissimilarity, entropy). For example, Itti et al. [1] first extracted the multi-scale center-surround contrasts from multiple features. These contrasts were then fused to estimate image saliency. Similarly, Gao et al. [5] computed image saliency as the discriminant power between center