Pattern Recognition 74 (2018) 359–372

Contents lists available at ScienceDirect

Pattern Recognition

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

Probabilistic saliency estimation

Caglar Aytekin*, Alexandros Iosifidis, Moncef Gabbouj

Department of Signal Processing, Tampere University of Technology, Tampere, Finland

ARTICLE INFO

Article history:
Received 30 January 2017
Revised 20 August 2017
Accepted 12 September 2017
Available online 20 September 2017

Keywords:
Saliency
Salient object detection
Spectral graph cut
Diffusion maps
Probabilistic model
One-class classification

ABSTRACT

In this paper, we model the salient object detection problem under a probabilistic framework encoding the boundary connectivity saliency cue and smoothness constraints into an optimization problem. We show that this problem has a closed form global optimum solution, which estimates the salient object. We further show that along with the probabilistic framework, the proposed method also enjoys a wide range of interpretations, i.e. graph cut, diffusion maps and one-class classification. With an analysis according to these interpretations, we also find that our proposed method provides approximations to the global optimum to another criterion that integrates local/global contrast and large area saliency cues. The proposed unsupervised approach achieves mostly leading performance compared to the state-of-the-art unsupervised algorithms over a large set of salient object detection datasets including around 17k images for several evaluation metrics. Furthermore, the computational complexity of the proposed method is favorable/comparable to many state-of-the-art unsupervised techniques.

1. Introduction

Salient object detection is a computer vision topic with growing interest over the last decade. The goal is to highlight the visually interesting regions in a given scene. The problem of saliency detection has been motivated by related work in neuroscience; humans select important visual information based on attention mechanisms in the brain [1]. Given this motivation, earlier works on saliency detection concentrate more on predicting sparse human eye-gaze points that are detected by eye-trackers [2]. Accordingly, most of the research on this track is based on biologically inspired algorithms, which try to imitate the dynamics of the human attention mechanism [2–9]. During the last few years another related track emerged where the goal is to segment *salient objects* [10–20], instead of predicting some sparse eye-fixations. Both research tracks produce saliency maps that are useful for tasks such as video surveillance [21], compression [22], image manipulation [23], automatic image cropping [24], foreground detection [25], and coding [26]. However, the output of salient object detector techniques is more useful, when compared to eye fixation predictions, for higher level computer vision and pattern recognition tasks such as tracking [27], object region proposals [29] and object recognition [28].

In this paper, we focus on the salient object detection task. Since ultimately we consider salient object detection as a preprocessing block for higher-level tasks as mentioned above, fast and generic methods would be preferred. Therefore, in this work we focus on unsupervised salient object detection. While supervised approaches, such as those in [42,33], have the potential of finding more accurate results, their performance depends on the training process followed and the data that has been exploited for training. Recent works have also indicated that unsupervised saliency detection approaches can compete (or even outperform) supervised methods [40].

Unsupervised salient object detection methods can be categorized based on the saliency cues they use. Commonly exploited cues include local and global contrast, boundary connectivity, shape and location cues. The local contrast cue is based on the assumption that the salient object is in contrast with its immediate surroundings [11,12,15,20]. A spectral foreground detection method was proposed in [20], which optimizes a criterion involving the minimization of cut-value, which is equivalent to the maximization of the local contrast. A region contrast based method was proposed in [11], which computes a salient region as a weighted sum of local contrasts with its surroundings. The global contrast cue is similarly defined as by assuming that the salient object is in high contrast with its surrounding. A histogram-based method enhancing regions with global contrast to the rest of the image was proposed in [11]. The boundary connectivity prior is one of the most widely used cues and is based on the assumption that most of the image boundaries will not contain parts of the salient object [10,15,16,20,36]. There are many ways to use this prior. For example in [20], the boundary pixels were strongly assigned to background in a belief vector, which was a part of the optimization problem

* Corresponding author.
E-mail addresses: caglar.aytekin@tut.fi (C. Aytekin), alexandros.iosifidis@tut.fi (A. Iosifidis), moncef.gabbouj@tut.fi (M. Gabbouj).

https://doi.org/10.1016/j.patcog.2017.09.023
0031-3203/

used to extract salient regions. A robust background assignment on image boundaries was conducted in [15] based on a hand-crafted measure that differentiates foreground and background regions touching the boundary. The boundary regions were used as background templates in [10] to re-construct the image in a sparse and dense way. Salient regions were defined as the ones that have high reconstruction errors [10]. In [16], boundary regions were defined as absorbing nodes in a Markov chain model. The saliency values of each region were evaluated based on the absorption time of the regions by the boundary nodes. Similar to the boundary connectivity cue, the center prior also assumes that salient objects are less likely to touch the image boundaries. However, it makes a stronger assumption that the object is mostly located in the center of the image. Although this is not a very reliable cue, it was used in [12] as a weighting coefficient on local contrast based saliency maps.

Besides the saliency cues, there are other supplementary processes that are widely used, such as using multiple resolutions [10,12,20] and smoothness constraints on saliency maps within similar regions [15,20].

The salient object detection methods can also be categorized according to their interpretation of saliency. One of the most popular interpretations is modeling saliency as a diffusion process [14,16,36]. These methods assume a preliminary information on saliency and designs several diffusion matrices to propagate this initial information to the entire salient region. Another interpretation of saliency is a graph-cut based one, where the salient object is considered as a foreground partition which constitutes a large area and a high contrast with the background partition [19,20,29]. The concepts behind these interpretations, i.e. diffusion maps and graph partitioning were shown to be related in an earlier work [35].

For an extensive survey on the recent state-of-the-art salient object detection algorithms, readers are encouraged to review [34]. Furthermore, an extensive benchmark comparing the performance of the state-of-the-art in salient object detection is also provided in [40].

In this paper, we model the salient object detection problem by following a probabilistic approach. We propose an unsupervised salient object detection method (which we call Probabilistic Saliency Estimation – PSE) that jointly optimizes saliency cues such as boundary connectivity and smoothness constraints. The solution of the proposed optimization problem is shown to have a closed-form. Moreover, we show that, along with the proposed probabilistic framework, this solution can also be interpreted from several perspectives, including spectral graph cut, diffusion and one-class classification approaches, all leading to the exact same solution. By further exploiting these links, we show that some diffusion based [14,16,36] and spectral graph based [20] salient object detection methods can also be cast into the proposed probabilistic framework, which provides us a theoretical platform for comparison of our method with them. Based on the above analysis provided in this paper, we also show that the proposed PSE method differentiates from our previous methods [19,20] from three aspects. First, it introduces a new formulation for saliency detection in a probabilistic framework. Second, by exploiting the graph-cut interpretation of PSE, we show that it also provides another approximation of the optimization criterion exploited in [19,20] with looser constraints. Finally, we show that [19,20] are suboptimal when they are investigated in the proposed probabilistic framework.

The main contributions of the paper can be listed as follows:

- A novel probabilistic approach to the unsupervised salient object detection problem is proposed. Following this approach, we formulate a novel optimization problem and we show that it has a global optimal solution (PSE), leading to state-of-the-art performance for unsupervised salient object detection. (Section 2)
- Based on the links observed between the proposed solution and those of diffusion based salient object detection methods, we interpret PSE as a diffusion based method. (Section 2.2)
- Based on the analysis of diffusion interpretation of PSE, we show that its solution involves terms related to graph cut. Therefore, PSE can also be interpreted as a solution to the graph-based cut problem. In fact, we show that PSE can be derived directly from the original graph based cut problem. (Section 2.3)
- We show that some diffusion based and graph based methods can be cast into our proposed probabilistic framework under some assumptions. (Appendices A and B)
- Finally, we show that the saliency detection problem can be interpreted as a one-class classification problem, whose solution is also given by PSE. (Section 2.4)

The rest of the manuscript is organized as follows. In Section 2, we introduce the proposed PSE method and explore its connections and differences to a wide range of salient object detection methods, namely the diffusion, the graph-based, and the one-class classification approaches. In Section 3, we present extensive experiments conducted on publically available datasets and compare the performance of the proposed method with the state-of-the-art unsupervised methods. Finally, we conclude the paper in Section 4.

2. Probabilistic saliency estimation

We formulate the salient object detection problem as that of the estimation of the probability mass function (PMF) $\boldsymbol{P}(\mathrm{x})$ of a random variable x, where a region/pixel x_i in an image is a possible outcome of the event. Since the summation over the PMF is constrained to be equal to 1, such a formulation allows us to set the total attention given to the scene as fixed which is analogous to the limited processing capability of the human brain. Accordingly, a study on the mechanisms of visual attention in the human cortex [1] states that: "*A typical scene contains many different objects that, because of the limited processing capacity of the visual system, compete for neural representation*". The probabilistic framework allows us to model such competition by defining an event as drawing distinct regions from the image, where $\boldsymbol{P}(\mathrm{x})$ defines the probability mass function of this event such that $\boldsymbol{P}(\mathrm{x} = x_i)$, probability of drawing the region x_i, is high if the region x_i is salient. This should not be confused with the probabilistic event of x_i being salient or not. Here, the event is selecting x_i among others as salient, introducing a competition between the regions. Within this framework, we wish $\boldsymbol{P}(\mathrm{x})$ to satisfy some properties related to the widely used saliency assumptions. For example, we expect similar probabilities for the image regions that are similar in a feature space, such as a color space. This is in accordance with the smoothness constraints for saliency as explained above. Furthermore, we wish to design $\boldsymbol{P}(\mathrm{x})$ in such a way that any prior information about the PMF can be considered. For example, in this way one can make use of the widely accepted assumption that image boundaries are more likely to belong to the non-salient region. Therefore, the estimation of $\boldsymbol{P}(\mathrm{x})$ can be formulated as an optimization problem that involves two terms: one that enforces similar regions in a given feature space to have similar probabilities and another that encodes any prior information about the PMF. We express this joint optimization as follows:

$$\underset{\boldsymbol{P}(\mathrm{x})}{argmin} \left(\sum_i (\boldsymbol{P}(\mathrm{x}=x_i))^2 v_i + \frac{1}{2} \sum_{i,j} \left(\boldsymbol{P}(\mathrm{x}=x_i) - \boldsymbol{P}\left(\mathrm{x}=x_j\right) \right)^2 w_{i,j} \right) \quad s.t. \quad \sum_{\boldsymbol{i}} \boldsymbol{P}(\mathrm{x}=x_i) = 1 \; . \tag{1}$$