Neurocomputing 205 (2016) 301–310

Contents lists available at ScienceDirect

Neurocomputing

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

Predicting salient object via multi-level features

Mingqiang Lin, Chenbin Zhang, Zonghai Chen *

Department of Automation, University of Science and Technology of China, Hefei 230027, PR China

ARTICLE INFO

Article history:
Received 25 November 2015
Received in revised form
13 February 2016
Accepted 6 April 2016
Communicated by Junwei Han.
Available online 10 May 2016

Keywords:
Saliency detection
Convex hull
Otsu
Principal component analysis

ABSTRACT

A wide variety of methods have been developed to predict where people look in natural scenes focused on pixel-level image attributes. Most existing methods measure the saliency of a pixel or region based on its contrast within a local context or the entire image. In this paper, we propose a novel salient object detection algorithm by integrating multi-level features including local contrast, global contrast, and background priors which measure the visual saliency in pixel-level, region-level, and object-level. We use the low level visual cues based on the convex hull to separate salient object from the background. The background priors are computed from the background templates using Principal Component Analysis. In order to suppress background noise, local and global contrasts are refined by object center priors which are computed with the Gaussian model based on background priors. Experimental results on four widely used public benchmark datasets demonstrate the proposed method performs well when against fifteen state-of-the-art methods in terms of precision and recall. We also demonstrate Otsu adaptive threshold method can be used to create high quality segmentation masks.

1. Introduction

Humans have a tremendous ability to rapidly direct their gaze when looking at a natural scene, and to select visual information of interest. The mechanism has proven to be useful for human as well as computer vision. Saliency detection can be applied to various computer vision tasks such as image segmentation [1], object recognition [2], adaptive compression of images [3], video object segmentation [36–38], content-aware image editing [4], image retrieval [5], object detection [6,44] and object tracking [7].

According to their mechanisms of representing image saliency, existing work can be roughly divided into two categories: bottom-up and top-down approaches. The bottom-up methods [8–10,14] are data-driven, and focused on integrating low-level features, such as contrast, location and texture. In the early works, Itti et al. [8] define local contrast using central-surround differences of image features. Cheng et al. [9] extend the histogram to 3D color space, and propose the global region contrast with respect to the entire image. Recently, Jiang et al. [14] formulate saliency detection as a semi-supervised clustering problem and use the well-studied facility location model to extract cluster centers for salient regions. Inspired by advances in compressive sensing research, Shen et al. [10] utilize low-rank and sparse matrix decomposition methods and their extensions for saliency detection. In contrast, the top-down methods [11–13] are often task-driven. The “feature-saliency” mapping is mainly guided by high-level priors. For example, Judd et al. [11] use a linear support vector machine to train a model of saliency. Yang et al. [12] use dictionary learning to extract region features and CRF to generate a saliency map. Liu et al. [13] learn a guidance map to fuse human prior knowledge to the LESD formulation for saliency diffusion. Han et al. [43] integrate appearance rarity with objectness likelihood in a probabilistic paradigm based on sparse coding representations.

On the basis of the model simplifying, the researchers found that contrast is the most important factor which dominantly influences human visual attention [9,15]. By defining pixel/region contrast in either local or global context, existing methods can be classified to two streams. Local methods [8,15,16] rely on pixel/region difference in the vicinity, while global methods [9,17,18] rely mainly on color uniqueness in terms of global statistics. Local methods tend to produce higher saliency values near edges instead of uniformly highlighting salient objects. Global methods can produce uniformly highlighting salient regions. However, global methods ignore spatial relationships across image parts, and may highlight background regions as salient.

The goal of salient object detection is to segment out background regions and thereby salient objects. Therefore, it needs to calculate the contrast between the objects and the image background and then select those with high contrast as salient objects. The local and global contrast based methods blindly assume the neighboring regions or the entire image to be the background,

* Corresponding author. Tel.: +86 55163606104.
E-mail address: chenzh@ustc.edu.cn (Z. Chen).

http://dx.doi.org/10.1016/j.neucom.2016.04.036
0925-2312/

which in turn reduces the performance of saliency detection. To overcome these problems, a few methods focus on background priors. Background priors aim to calculate the contrast between the objects and the image background. Several recent approaches [19–22,41,42] exploit boundary priors to generate background priors. Such methods assume the image boundary is background, suggesting that boundary priors are effective. However, boundary assumption is fragile and may fail even when the object only slightly touches the boundary. The work in [22,42] also observe this drawback. Zhu et al. [22] define the boundary connectivity as a Robust Background Measure. Han et al. [42] formulate the measure of background priors as the reconstruction residuals and then use the center of the salient cluster to refine saliency map. Related to our background priors is the work of Liu et al. [40] which computes the prior map based on the results of the superpixel clustering. However, the saliency of a segment is simply measured by the number of pixels belonging to the convex hull. Jiang et al. [41] combine the regional contrast, regional property and regional backgroundness descriptors together to form the master saliency map, but the pseudo-background region is defined as the 15-pixel wide narrow border region of the image.

Aiming to solve this notorious and universal problem, we use multi-level features including local contrast, global contrast, and background priors which measure the visual saliency in pixel-level, region-level, and object-level. Computing local contrast based on pixels can better retain the boundary information of salient object. Computing global contrast based on regions can preserve relevant structure information, and ignore unnecessary texture information. The boundary priors work better for off-center objects but are still fragile and can fail even when an object only slightly touches the boundary. Hence, we use the low level visual cues based on the convex hull to separate salient object from the background. The background priors of object are computed from the background templates using Principal Component Analysis (PCA). In order to suppress background noise, local and global contrasts are refined by object center priors which are computed with the Gaussian model based on background priors. In general, methods that utilize high-level information to obtain more informative saliency priors perform better than purely low-level approaches. We have extensively evaluated our methods on publicly available benchmark data sets, and compared our method with state-of-the-art saliency methods. The experiments show significant improvements over previous methods both in precision and recall rates. We also demonstrate Otsu adaptive threshold method can be used to create high quality segmentation masks.

2. Overview

As motivated before, we propose an algorithm that first compute multi-level features including local contrast, global contrast, and background priors. Secondly compute object center priors with the Gaussian model based on background priors. Then refine local and global contrasts by object center priors. Finally effectively combine these refined contrasts and background priors for salient region detection. Hence, our algorithm consists of the following steps (see Fig. 1).

2.1. Contrast

Based on the observation from biological vision that the vision system is sensitive to contrast in visual signal, we propose a set of contrasts including local contrast and global contrast. Local contrast measures the “rarity” of each pixel in the vicinity. Global contrast describes the spatial distribution of a specific color.

2.2. Background priors

Recent works introduce boundary priors and treat image boundary regions as background. This assumption is fragile and may fail even when the object only slightly touches the boundary. Considering the drawback of boundary priors, we use the low level visual cues based on the convex hull to separate salient object from the background. The background priors of object are computed from the background templates using PCA.

2.3. Object center priors

Some early works [25–27] use the center priors to bias the image center region with higher saliency. However salient objects do not always appear at the image center. Therefore, the proposed object center priors are computed with the Gaussian model based on background priors.

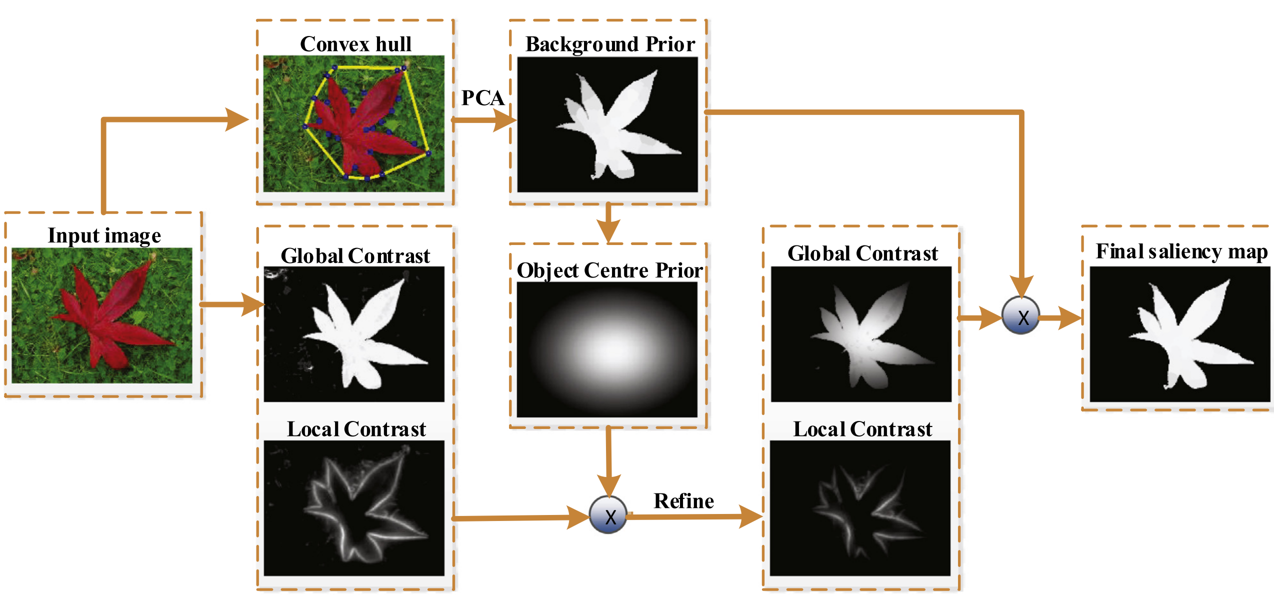

Fig. 1. Illustration of the main phases of our algorithm.