Neurocomputing 194 (2016) 348–359

Contents lists available at ScienceDirect

Neurocomputing

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

Local graph regularized sparse reconstruction for salient object detection

Lina Huo *, Shuyuan Yang *, Licheng Jiao *, Shigang Wang, Shuang Wang

Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, Xidian University, Xi'an 710071, China

A R T I C L E I N F O

Article history:
Received 13 May 2015
Received in revised form
17 September 2015
Accepted 15 February 2016
Available online 4 March 2016

Keywords:
Graph Laplacian
Graph sparse reconstruction
Local similarity
Salient object detection

A B S T R A C T

Subspace representation based salient object detection has received increasing interests in recent years. However, due to the independent coding process of sparse reconstruction, the locality and the similarity among regions to be encoded are not explored. To preserve the locality and similarity of regions, a graph Laplacian regularization term is constructed as a smooth operator to alleviate the instability of the salient score in visual object. Then a new saliency map is calculated by incorporating this local graph regularizer into sparse reconstruction, which explicitly explores the local spatial structure of salient objects and thus obtains more uniform salient map. Moreover, we advance a heuristic object based dictionary from background superpixels, by which objects can be more accurately located. Experimental results on four large benchmark databases demonstrate that the proposed method performs favorably against fifteen recent state-of-the-art methods in terms of five evaluation criterions.

1. Introduction

Human are able to rapidly and effortlessly select the important information (so called salient) from the large amount of received data and perceive the scene. For simulating such capability in machine, saliency detection is generally considered as a selection process of the interesting regions that attract the visual attention in a scene. In the past decades, a great number of saliency detection models have been proposed and considerable progress have been made. It is originally used in eye fixation prediction [1,2] to investigate where people look during free-viewing of static nature scenes. Whereas recent extended trend concentrates on salient object detection [3–5] to accurately identify the salient region in all scenarios and then segment the accurate boundary of salient objects. Recently, researchers have shown an increasing interest in automatic saliency detection because it can help us to find the interesting objects or regions to effectively represent a scene. It has served as an important preprocessing procedure for a wide range of computer vision applications such as object classification and detection, image segmentation, active vision, scene understanding, etc.

The salient object detection is commonly interpreted as a process that estimates the probability of per-pixel belonging to salient objects and accurately segments the salient foreground objects from the background. One of the earliest saliency models of Itti et al. [1] implemented the bottom-up attention based on center-surrounded mechanisms across multi-scale image features to pay attention to the region. A large number of methods have been proposed to extend this method, which combine local, regional and global contrast-based features to define the generic salient object, such as the fuzzy growing method [2], and graph-based approach [6], information maximization [7]. Subsequent works of Liu et al. [3] and Achanta et al. [4] defined saliency detection as a segmentation problem which detected the most salient object and segment the accurate boundary of the object. Recently, a new trend was to formulate the problem of saliency detection with the framework of subspace segmentation, in which salient object, in terms of uniqueness, can be defined as sparse coding in a certain feature space. Shen and Wu [8] proposed that the image can be decomposed as a low-rank matrix plus a sparse noises matrix in a certain learned feature transformation space, where the background region corresponded to the low-rank matrix, and the salient regions were indicated by the sparse noises. Zuo et al. [9] proposed a bottom-up segmentation as a guidance cue of the matrix recovery framework which used the self-representation idea to compute saliency scores. In the sparse appearance model [10], the saliency measures for the other regions are less accurate due to inaccurate inclusion of foreground segments as part of sparse basis functions.

In classical sparse coding scheme [10], salient objects, in terms of uniqueness, can also be defined as the *sparse noises* in a certain feature space. This sounds reasonable and remarkable experimental results

* Corresponding authors. Tel.: +86 2988202661; fax: +86 2988201023.
E-mail addresses: huolinalw@163.com (L. Huo), syyang@xidian.edu.cn (S. Yang), lchjiao@mail.xidian.edu.cn (L. Jiao), sgwang@mail.xidian.edu.cn (S. Wang), shwang@mail.xidian.edu.cn (S. Wang).

http://dx.doi.org/10.1016/j.neucom.2016.02.044
0925-2312/

have been demonstrated. However, directly using sparse reconstruction in feature space of images still encounters some difficulties for the task of meaningful and reliable saliency detection. On the one hand, the coefficients of sparse representation are less stable, which may lead to non-uniform saliency detection result of similar regions. In other words, similar regions may be encoded as totally different sparse coefficients of the background templates. Sparse reconstruction methods independently compute the saliency of each image region and ignore the interrelationships among the adjacent regions (e.g., similar regions should have similar saliency), which can be critical for uniformly highlighting the whole salient region and adequately suppressing the background region. On the other hand, when the dictionary contains some foreground regions, their saliency measures are low sparse reconstruction errors due to the sensitiveness of the dictionary. It assumed that the salient object is often framed near the center of image. However, salient objects often appear in the image boundary in many images. When foreground segments are mistakenly included in the background templates, solutions (i.e., coefficients) by sparse representation are falsely mark background as the salient regions should be low.

To alleviate the impact of these problems, we introduce a local graph regularized sparse reconstruction algorithm which explicitly preserves geometrical structure of regions to alleviate the instability of representations. The graph Laplacian term as a locality preserving term is considered the locality and similarity of regions in the image. The fidelity term is incorporated into the graph Laplacian regularized term to obtain smooth representations so as to uniformly highlight the entire salient object. Since some foreground segments as part of sparse basis functions may lead to discontinuous saliency detection results, we propose a heuristic background dictionary based on image boundary information which removes the foreground noises from the border regions. Our method can have much more discriminating power and effective representation than the traditional sparse coding algorithms. The experimental results show the effectiveness on the four popular datasets covering various foreground statistics.

The rest of this paper is organized as follows: we review the related works on salient object detection in Section 2. In Section 3, we provide a brief description of local graph regularized sparse reconstruction model and a classical algorithm for feature-sign search algorithm [28] to optimize this problem. Section 4 demonstrates that our model performs better in detecting salient objects than the state-of-the-art approaches in four popular datasets.

2. Related works

In the past decades, detecting salient or interesting objects in natural scenes has attracted a lot of focused research in computer vision. Visual saliency is typically measured by image contrast computation over different features such as intensity, color, or texture. A comprehensive review of saliency detection and visual attention can be seen in [12,13], and a qualitative and quantitative analysis of different methods was provided in [24]. According to the extent of context in which the contrast of each image region was calculated, bottom-up salient region detection methods can be broadly classified into local-contrast based methods and global-contrast based methods.

Local contrast based methods estimate the appearance uniqueness of each image pixel/region with respect to small local neighborhoods. Motivated by mechanisms of the human visual search strategies, Itti et al. [1] presented the center-surrounded differences across multi-scale image features to pay an attention to the region. A fuzzy growing process was proposed to simulate human perception mechanism [2], by which the attended points, attended areas and attended view were directly extracted. Some local representative methods calculated the saliency map of the input image in different frameworks, such as graph-based approach [6], center-surround divergence [14] and incremental sparse coding [15]. While these approaches aimed to identify a certain region with high visual stimuli, they usually tended to highlight the object boundaries and small objects in images. However, due to their overly-emphasized local difference, purely local contrast models failed to detect the inner regions of the object, limiting their applicability for some vision tasks such as image segmentation, image compression and object detection.

Global contrast based methods evaluate the saliency of an object as the uniqueness in the entire image. In contrast to local methods mentioned above, global contrast can get more uniform saliency regions by computing the dissimilarities among all pixels or regions in the entire image. Achanta et al. [16] proposed a frequency-tuned approach to compute full resolution saliency maps in images using low level features of color and luminance. This approach, however, only considered the pixel-wise color difference between the smoothed image pixels and the average color of the image, which can be insufficient to analyze complex variations common in natural images. Cheng et al. [5] introduced a regional contrast based salient object detection algorithm, which considered spatial relationships across image regions to obtain coherence saliency scores. Intuitively, a region is considered to be salient if it is remarkably distinct from its most similar regions, while their spatial distances are taken into account. Margolin et al. [17] defined the uniqueness of a patch by measuring its distance to the average patch based on the observation that non-distinct patches were more concentrated than distinct ones in the high-dimensional space. Some studies of visual saliency detection methods selected the salient regions of the input image in the transformed domain [18–21]. Hou et al. [18] built a visual saliency model by extracting the spectral residual of an image in the spectral domain, and claimed that the spectral residual model can be implemented by log-spectrum representation of images. Wei Wang et al. [20] defined saliency as site entropy rate (SER) based on information maximization principle, which measured the average information transmitted from a node to all the others during the random walk on graphs or networks. Xie et al. [19] integrated the smoothing constraints into the framework of sparse coding to group superpixels in the image, and mid-level cues originating from varying superpixel size were also taken into consideration. A weighted sparse coding of saliency detection framework [21] was presented to handle heterogenous types of input image. In classical sparse coding scheme [10], the sparse reconstruction error was defined as the residual based on the sparse representation of the background templates. As opposed to local contrast methods, global contrast approaches can obtain more uniform salient regions. Furthermore, these methods still cannot uniformly highlight the entire object effectively, due to ignorance spatial relationship across image regions. In order to achieve better performance, we add the graph Laplacian term as a locality preserving term to highlight more uniformly salient object regions.

Essentially, the true aim of salient object detection is to detect and segment the most salient objects that are distinctive from the image background. Recently, some studies of visual saliency detection methods often use background contrast to select the salient regions of the input image, assuming that regions along the image boundary are more likely to be the background [22–25]. In [22], a robust background measure, called boundary connectivity, was exploited to formulate the background region was heavily connected to the image boundary. And based on graph-based manifold ranking, the work of Yang et al. [23] utilized the four boundaries of the input image as background prior to extract foreground queries for the final saliency map. In classical sparse coding scheme [10], the regions along the image borders were extracted to construct the background dictionary. Saliency was defined as the residual based on the sparse representation of the