Journal of Visual Communication and Image Representation 50 (2018) 270–279

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

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

Salient object detection via spectral graph weighted low rank matrix recovery☆,☆☆

Jiazhong Chen[a,*], Jie Chen[b], Hefei Ling[a], Hua Cao[a], Weiping Sun[a], Yebin Fan[a], Weimin Wu[c]

[a] School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China
[b] International School, Jinan University, Guangzhou 510660, China
[c] Department of Information Technology, Fujian Chuanzheng Communications College, Fuzhou 350001, China

ARTICLE INFO

Keywords:
Saliency detection
Spectral graph
Low rank matrix recovery
Sparse decomposition
Feature matrix

ABSTRACT

A novel saliency detection method via spectral graph (SG) weighted low rank matrix recovery (LR) is presented in this paper. The location, color, and boundary priors are exploited in many LR-based saliency detection methods. However, these priors do not work well when the salient objects are far away from image center, especially when the background is complicated and has low contrast with objects. Because spectral graph contains rich image contrast, it is used as an efficient weight to obtain a much reasonable high-level prior in the proposed LR-based saliency model. Compared with previous LR-based methods, low rank matrix and sparse matrix rather than only sparse matrix are used to calculate the final saliency by an integration function and an activation function. The numerical and visual results on four challenging salient object datasets show that our method performs competitively for salient object detection task against some recent state-of-the-art algorithms.

1. Introduction

Saliency detection devotes to identify visually attention regions or objects in images to enhance existing computer vision system such as image retrieval [1], automatic object detection [2], image retarget [3], video summarization [4], and adaptive image compression [5]. It becomes an interesting and challenging research area in recent years and involves a wide range of knowledge from cross-disciplines like neurology, biology, computer science, machine learning, mathematics, and so on.

The LR is introduced successfully in two directions of saliency detection: human eye fixation detection and salient object detection. In the first direction, Yan et al. propose an early approach towards visual fixation detection via matrix decomposition [6], and Lang et al. proposed a multi-task sparse pursuit method for detecting saliency [7]. In the second direction, Candès et al. introduce an early approach towards salient object detection via low rank matrix recovery [8]. Later, Shen et al. proposed a unified method based on LR to incorporate traditional low-level features with high-level prior knowledge [9].

The background may be complicated and have low contrast with objects in real images, but the classical LR-based models neglect the underlying structure of images. This denotes directly applying the LR model to the saliency detection has limited robustness and inevitably degrades the associated performance. To cope with this issue, Zou et al. exploit bottom-up segmentation as a guidance cue of the matrix recovery [10], and Peng et al. propose a background prior weighted structured sparse matrix decomposition model [11]. Moreover, in some cases, the salient objects with big size occupy most of the image and do not satisfy the sparse assumption. To address this problem, a high-level background prior is estimated by employing the location, color, and boundary priors to weight the image feature matrix in [12]. However, the problem caused by big object size is still not addressed very well. Moreover, these priors still do not work well when the salient objects are far away from image center, especially when the background is cluttered and has low contrast with objects. Because spectral graph contains rich image contrast [13], it is introduced as an extra prior in this work to deal with these issues caused by object size and location, background complexity, and low contrast between background and objects.

Thus a new spectral graph weighted low rank matrix recovery (SGLR)-based salient object detection method is presented in this paper. Because the SGLR can suppress the image background in the low rank matrix and sparse matrix much efficiently than previous LR methods, both of them are used for final saliency generation by an integration

☆ This paper has been recommended for acceptance by Zicheng Liu.
☆☆ This work was supported in part by the Natural Science Foundation of China under Grant U1536203 and Grant 61300140, and in part by the Major Scientific and Technological Innovation Project of Hubei Province under Grant 2015AAA013.
* Corresponding author.
E-mail addresses: chenjz70@163.com, jzchen@hust.edu.cn (J. Chen).

https://doi.org/10.1016/j.jvcir.2017.12.006
Received 6 March 2017; Received in revised form 9 December 2017; Accepted 16 December 2017
Available online 16 December 2017
1047-3203/

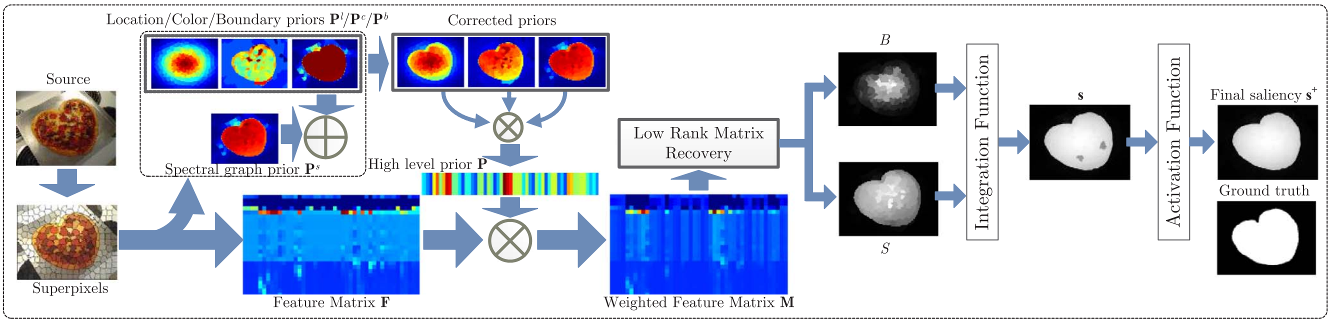

Fig. 1. Diagram of proposed saliency detection method.

function in our work.

The framework of our SGLR-based salient object detection is shown in Fig. 1. The main contributions of this work are fourfold: (1) an effective spectral graph prior is employed to correct the existing priors for obtaining a reasonable high level prior, (2) a solver of SGLR decomposition is presented for the image feature matrix weighted by the high level prior, (3) compared with previous LR-based methods that only use sparse matrix for saliency estimation, because the high performance of matrix decomposition is achieved by the proposed SGLR model, the low rank matrix and sparse matrix, which are the outputs of the SGLR decomposition, are both used in final saliency calculation, and (4) an efficient integration function with an activation function is presented to fuse the low rank matrix and the sparse matrix for final saliency generation. The numerical result shows that our model performs competitively for salient object detection task against some recent state-of-the-art algorithms. The visual result shows our model tackles the above mentioned issues very well.

The rest of the paper is organized as follows: Section 2 presents a review of related work. Section 3 provides a scheme to generate a reasonable high level prior from spectral graph for weighting image feature matrix. Section 4 describes a solver of LR decomposition for spectral graph weighted feature matrix, and a detailed numerical and visual evaluation for the proposed method is presented in Section 5. The conclusions are given in Section 6.

2. Related work

According to the motivation and the technical components of this paper, especially the high level visual concepts, the related work is reviewed in three aspects: (1) saliency detection, (2) saliency detection via deep learning, and (3) saliency detection via LR.

To estimate visual saliency, many approaches have been proposed in the past few decades, which can be categorized as either top-down [14,15] or bottom-up [16–26] approaches. All bottom-up saliency methods rely on prior knowledge about salient regions and backgrounds. Different saliency methods characterize the prior knowledge from different perspectives. Most of the bottom-up methods can be categorized into two basic classes depending on the definition way of saliency cues: contrast prior and background prior [16].

Saliency detection is originally for predicting human eye fixation. For example, Itti et al. extract center-surround contrast at multiple spatial scales to find the prominent regions [17], and Parkhurst et al. use a purely bottom-up model of selective visual attention to examine the degree to which eye movements are determined by stimulus properties alone [18]. Many follow-up works are along this direction [19–22]. Recent years have witnessed more interest in object level saliency detection. In this direction, the salient objects are automatically detected and assigned consistently high saliency values. For example, Cheng et al. incorporate spatial relations to produce patch-based contrast to measure saliency [23], Perazzi et al. show the saliency can be estimated by unifying two contrast measurements so-called uniqueness and spatial distribution of image elements [24], and Wei et al. exploit geodesic distance of inner and boundary patches to calculate the contrast [16].

The semi-supervised learning (SSL) framework has attracted an increasing amount of considerable attention in recent years [27–30]. The essence of SSL is to exploit the prior knowledge from the unlabeled data. Gopalakrishnan et al. formulate the object detection problem as an automatic labeling problem on a weighted graph [27]. The most salient seed and several background seeds are first detected using Markov random walks performed on two different graphs. However, it is challengeable to determine where and how many the seeds should be selected as their generation manner, especially for the scenes with more than one salient object. Please note although the methods in [27–30] perform the saliency detection in a SSL framework, the detection is undertaken in an unsupervised manner.

Assuming that most image boundary regions are background, the conception of connectivity and boundary priors is presented for salient object detection in [16]. Based on the boundary prior, the labels of boundary nodes are initialized strongly as "1" in [28]. However, this empirical manner can not provide the label fitness very well when the salient objects are partially cropped by image boundaries. To address this issue, Li et al. compute the color distinctiveness of each boundary node from other boundary regions and drop the top 30% with high color difference to obtain the background seeds, which are further assigned with "1" as their initial labels [29]. Zhang et al. set the pixels along the image boundary as the seeds, and compute the minimum barrier distance (MBD) transform for each color channel to capture the boundary connectivity cue. Then the MBD maps for all color channels are pixel-wise added together to form a combined MBD as saliency [31]. The authors of [32] use the distance between pixels on minimum spanning tree to measure the boundary dissimilarity as saliency. Instead of estimating saliency only from the features within a query image, Wang et al. transfer annotations from support images into the query image according to their global and local correspondences [33].

Saliency detection in video is helpful to image re-ranking and motion deblurring. For example, Huang et al. introduce visual attention into image re-ranking and develop a new model for detecting images which have salient object [34], and Zhang et al. propose a novel motion deblurring approach based on saliency detection [35]. For motion region segmentation, the motion cues of video objects provide indication for salient video regions [36]. Besides temporal motion cues, spatial edges are exploited as indicators of foreground object locations for salient video object segmentation [37]. Considering different foreground motion patterns, the gradient flow field is deeply incorporated with spatial edge and motion features to indicate the locations of salient areas [38]. In [39], the spatio-temporal SIFT (scale-invariant feature transform) features of video objects are explored in intra-frame saliency, inter-frame consistency, and across-video correspondence.

Deep learning has been introduced into saliency detection domain for several years. Compared to traditional methods, it can incorporate high level visual concepts into a saliency detection task. Han et al. use a