J. Vis. Commun. Image R. 32 (2015) 144–155

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Salient region detection through sparse reconstruction and graph-based ranking ☆

Mian Muhammad Sadiq Fareed, Gulnaz Ahmed, Qi Chun *

School of Electronic and Information Engineering, Xian Jiaotong University, Xian 710049, China

A R T I C L E I N F O

Article history:
Received 3 April 2015
Accepted 3 August 2015
Available online 7 August 2015

Keywords:
Super-pixels
Sparse saliency
Separation saliency
Graph-based ranking
Bayesian integration
Location prior
PR curve
F-measure

A B S T R A C T

In this paper, we propose a salient region detection algorithm from the point of view of unique and compact representation of individual image. In first step, the original image is segmented into super-pixels. In second step, the sparse representation measure and uniqueness of the features are computed. Then both are ranked on the basis of the background and foreground seeds respectively. Thirdly, a location prior map is used to enhance the foci of attention. We apply the Bayes procedure to integrate computed results to produce smooth and precise saliency map. We compare our proposed algorithm against the state-of-the-art saliency detection methods using four of the largest widely available standard databases, experimental results specify that the proposed algorithm outperforms. We also show that how the saliency map of the proposed method is used to discover outline of object, furthermore using this outline our method produce the saliency cut of the desired object.

1. Introduction

Saliency detection is a mechanism of capturing human attention and finding out the most informative and useful part in an image/video. It selects the most important information from a scene and filter out the remaining unwanted part. From last few decades the saliency detection is widely studied because it extracts the visually salient region which is closer to human visual system. Human visual system has many applications in computer vision such as image re-targeting [1,2], image segmentation [3,4,54,57], object detection, image adaptation, object recognition [5] and image retrieval [6]. Different approaches have been designed because each image has different contents and carries different information. We can categories these saliency detection approaches as bottom up [7,8] and top down [9,10,58]. The bottom up approach is data driven. It depends on the prior knowledge of an object and background. While the later one is task dependent which entails with supervised learning and class labels.

Visual saliency [41,42] is divided into two categories like low level and high level features extraction. In low level features extraction, saliency is considered based on the color, contrast and intensity of a region with respect to the remaining scene. In second case saliency is calculated using high level factors like human faces. Among above mention categories there are some methods which compute saliency locally [11], globally [12,13], local–global [14] and using the contrast based process [15–17]. Present methods calculate visual saliency in three steps [18]. In first step, low or high level features are extracted from the image/video. In second step, on the basis of these extracted cues the saliency is calculated. And in third step the computed saliency is represented after normalization.

Some algorithms for saliency detection are based on the *RGB* model. For equally salient region calculation the near-infrared region is also incorporated with RGB model [18]. Because the near-infrared region is proved to be more informative than conventionally three colors and provides more clues for recognition and categorization. In [19], the authors proposed a model which adopts low level features like color, luminance, texture and depth for saliency calculation. They converted RGB image to YC_bC_r color space due to its perceptual property.

The algorithm [20–24] used graph based technique to find the saliency detection in which they transform the direction of gaze from one object to another. The human eye movements are used in these models as a random walks on graph. They divide the image into K mean clusters and suppose each cluster as a node. The authors assigned a transitional probability to each edge of a graph which is proportional to dissimilarity between the features of two connected nodes. Then Markov chain is used to estimate the saliency of each node by stationary probability. However, the saliency

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author.
E-mail address: qichun@mail.xjtu.edu.cn (Q. Chun).

http://dx.doi.org/10.1016/j.jvcir.2015.08.002
1047-3203/

results are strongly affected by graph scale due to the difference between block size and image.

Optimal contrast based saliency detection [25] uses entropy rate to determine the size of the object. They first calculated the contrast using the weighted sparse coding residual algorithm and then computed the contrast hypothesis by varying the size of surroundings. After that they utilized this optimal contrast to find out the final saliency map. To achieve the good performance results they used the multi-scale saliency enhancement. The bottom up method [26] uses the background prior to compute the saliency. They exploit the boundary and border prior for their saliency results calculation. Authors rely on priors to measure contrast between background and object. Contrast prior mainly used to determine the difference between pixels or regions. They assumed that the image boundary is background and regions at the boundaries are connected to each other. They named these as boundary priors and connectivity prior. Based on the boundary prior and background prior they design their saliency map. However, they attenuate the object from the inner side and found proper boundaries of salient object.

In literature, a lot of work has been done on object at the center of an image. In [30], the authors investigate and explore that from human fixation point of view the objects are always near the center part of an image, they made this preconception a strong base for setting up a series of experiments and found that objects are always placed in center of the image because the photographer wants to capture the whole scene without losing any information. From their experimental results they also found that if we use a Gaussian blob which is placed at the center of the image it produces more accurate and fine results for saliency. In [31,32], a central bias method is introduced which assigns weight to center region to compute saliency. However, for many images in which the objects are not close to the center of the scene, this supposition is not very logical.

The receptive fields of simple cells in primary visual cortex yield a sparse representation to represent natural images by sparse features [59]. The dictionary or the basis for sparse representation does not stay fixed however, varies with stimulus and as a result better represents the current environment. Independent component analysis (ICA) is a very famous method for multi-linear data analysis. Basis functions provided by ICA are non-Gaussian as well as statistically independent. An image can be expressed using ICA dictionaries, and subsequently the coefficients of these dictionaries are similar, sparse to neural receptive fields in the visual cortex [60]. Sparse representation has been utilized in a range of applications including compression, regularization in inverse problems, feature extraction, object detection and de-noising [61].

Our contribution: In this work, we propose a salient region detection algorithm which aims to beat the above discussed issues in salient object detection methods. Perhaps all saliency computation approaches exploit a color channel. Several have utilized *RGB* color model [33] whereas, others have engaged Lab color space [34]. Motivated by these findings it is better to estimate human color sensitivity. We argue that to make use of just one color system does not lead all the times to successful detection. Our method shows that attractive objects in several images are more accurately salient in Lab color space, whereas, for various other saliency detection methods better in *RGB*. The designed method uses not only contrast prior for salient region detection, although incorporate two background priors, and demonstrate better performance compared with preceding contrast based methods as described in Fig. 1. The proposed model performs very well when the color of the salient object is similar to the background or when contrast is very less. However, MR [48] and DSR [29] results are not very good as described in third and fourth column respectively in Fig. 1. To produce more even and accurate saliency results for each super-pixel in an image Bayesian integration enhancement scheme is utilized. Quantitative visual comparison of our method with thirteen state-of-the-art methods and graphical comparison with twenty-two state-of-the-art schemes on human-annotated benchmark data-sets are also performed to check the validity of the proposed model. The major contributions of the proposed approach are summarized as follows:

1. We extract visual information by using background contrast from each image and assign each super-pixel with a unique saliency based on that extracted visual information. Then we use graph-based ranking scheme which implements the ranking procedure to arrange the salient super-pixels in the surrounding regions based on their assigned values.
2. We use the Gaussian blob with different values of image center coordinates and super-pixel coordinates to enhance saliency results.
3. The proposed model automatically detects the outlines of objects of the interest on large data-sets. Additionally, our method generates the precise saliency cut of the desired object.

The rest of this paper is organized as follows. We review the previous approaches for saliency detection in Section 2. In Section 3, particulars of the presented algorithm are illustrated to show how the improvements are achieved from pixel to region level to extract saliency in an image. In Section 4, the experimental results and comparisons are discussed to validate the proposed scheme. Finally, the conclusion is drawn in Section 5.

2. Related work

To get basic graph arrangement information and to include the local grouping indications in graph labeling [48], the authors used Manifold Ranking (MR) procedure to find out a ranking function which is learn from an optimal affinity matrix. The state-of-the-art saliency detection method through graph-based ranking needs information about only one feature to make graph. However, MR need boundary prior and contrast prior for saliency detection. Additionally, it is complicated to decide the number and positions of salient cues as they are produced by random walks, particularly for those images which have different salient objects. In MR, all the background and foreground quarries can be simply generated through background prior and saliency is ranked through these quarries. Consequently, if the background is complicated or contrast is very less the computed saliency is not so accurate as shown in Fig. 2. By using boundary prior a human fixation model is proposed in [35], which suppose that human foci of attention is near the center of the scene. The above mentioned cues and priors are also employed in [36] to extract salient object.

Dense and Sparse Reconstruction (DSR) for saliency detection is proposed in [29]. DSR first computes sparse and dense reconstruction errors. Then these reconstruction errors are propagated on the context obtained by K-means clusters. Secondly, DSR utilizes a multi-scale reconstruction error followed by object biased Gaussian process to get smooth region base saliency. Finally, these dense and sparse propagated reconstruction errors are fused using Bayesian integrations method to achieve saliency results. However, DSR does not achieve persuasive results when the contrast is very less. Sometimes it includes background pixels or add artifacts inside object so that the salient object is not clearly distinguished from background.

In [17], a theoretically clear and perceptive method for contrast based visual saliency computation is discussed. To avoid the unnecessary information and computation they divided the image into similar dense regions. Then they calculate the rarity