Digital Signal Processing 55 (2016) 22–31

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

A convex hull approach in conjunction with Gaussian mixture model for salient object detection

Navjot Singh [a,b,*], Rinki Arya [a], R.K. Agrawal [a]

[a] School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi 110067, India
[b] National Institute of Technology, Uttarakhand, Srinagar, Pauri (Garhwal) – 246174, India

ARTICLE INFO

Article history:
Available online 6 May 2016

Keywords:
Salient object detection
Keypoint detection
Convex hull
Gaussian mixture model
Expectation maximization
Saliency map

ABSTRACT

The capability of humans in distinguishing salient objects from background is at par excellence. The researchers are yet to develop a model that matches the detection accuracy as well as computation time taken by the humans. In this paper we attempted to improve the detection accuracy without capitalizing much of computation time. The model utilizes the fact that maximal amount of information is present at the corners and edges of an object in the image. Firstly the keypoints are extracted from the image by using multi-scale Harris and multi-scale Gabor functions. Then the image is roughly segmented into two regions: a salient region and a background region, by constructing a convex hull over these keypoints. Finally the pixels of the two regions are considered as samples to be drawn from a multivariate kernel function whose parameters are estimated using expectation maximization algorithm, to yield a saliency map. The performance of the proposed model is evaluated in terms of precision, recall, F-measure, area under curve and computation time using six publicly available image datasets. Experimental results demonstrate that the proposed model outperformed the existing state-of-the-art methods in terms of recall, F-measure and area under curve on all the six datasets, and precision on four datasets. The proposed method also takes comparatively less computation time in comparison to many existing methods.

1. Introduction

Salient object detection [1] is a fundamental and significant research problem in the field of psychophysics, neurophysiology and their computational modeling perspectives [2]. It tries to imitate the human visual system by focusing on regions of interest present in a complex scene. The region of interest contains an object of a specific category which is unknown a priori but is dominant in an image. Human vision utilizes the visual attention mechanism to detect these dominant objects, popularly known as salient objects. Salient object detection finds applications in surveillance systems [3], remote sensing [4] and image retrieval [5,6]. It is helpful in automatic target detection [7,8], robotics, image and video compression [8], automatic cropping/centering [9] to display objects on small portable screens [10], medical imaging [11], advertising a design [8], image collection browsing [12], image enhancement [13] and many more.

Visual attention [1,14] can be achieved by using bottom-up and/or top-down approaches. Bottom-up approaches are fast, stimulus driven and task independent. They extract certain low-level features from the image and combine them into a saliency map [7]. The features can be extracted either at the local level or global level. While top-down attention is driven by cognitive factors such as knowledge, expectations and current goals. They are slow and task dependent. Top-down approaches are integrated with the bottom-up approaches in order to detect the salient locations. Most of research works mostly focused on the bottom-up aspect of visual attention. With the advancement of these bottom-up approaches, researchers started distinguishing the two very similar terms: fixation prediction and salient object detection. The fixation prediction models try to mimic the human vision with an objective that the human eyes mainly focus on some of the points in a given scene if shown for a few seconds. These points are helpful in eye movement prediction. The second category of models which are salient object detection models detects the most salient object in an image by segmenting the image into two regions, a salient object and background, by drawing accurate silhouettes of the salient object. Both categories of models construct saliency maps which are useful for different purposes.

* Corresponding author at: School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi 110067, India.
E-mail addresses: navjot.singh.09@gmail.com (N. Singh), rinki.arya89@gmail.com (R. Arya), rkajnu@gmail.com (R.K. Agrawal).

http://dx.doi.org/10.1016/j.dsp.2016.05.003

It all started about two decades ago, when Itti et al. [7] (Itti) implemented the model which utilized feature integration theory to combine intensity, color and orientation feature maps into a saliency map and gave the first fixation prediction model as suggested by Christof Koch [15]. Later Bruce and Tsotsos [16] (AIM) modeled visual saliency by using the concept of information maximization. Han et al. [17] extended the Itti et al. [7] model by using region growing techniques. Meur et al. [18] computed the saliency for the chromatic as well as the achromatic channels by using subband decomposition. Harel et al. [19] (GBVS) proposed a novel graph based visual saliency model. Hou and Zhang [20] (SR) gave a simple and fast model in terms of the spectral residual of the image. Later they have claimed that the SR theory is wrong and the right explanation for the remarkable performance of SR method can be found in their subsequent paper on image signature [21]. Liu et al. [22] (Liu) extracted features at the local, regional and the global level and used a supervised approach to partition the image into attention region and background region. Later in 2011 they extended their work on videos as well [23]. Cheng et al. [24] utilized the concept of global contrast differences and spatial coherence to detect salient objects. Yu and Wong [25] proposed a grid cell based image segmentation algorithm to extract salient objects. Zhang et al. [26] (SUN) evaluated the probability of a target at every location in the image based on Bayesian framework to determine salient object. Achanta et al. [27] (FT) generated a frequency tuned saliency model by using an image subtraction technique. Achanta and Susstrunk [28] (ASS) utilized the concept of maximum symmetric surround difference to yield a saliency map. Zhang et al. [29] used a combination of position, area and intensity saliency, and Bayesian framework to classify a pixel into an attention pixel or a background pixel. Chang et al. [30] proposed a model for salient object detection based on the fusion of visual saliency and generic objectness. Goferman et al. [31] (Gof) proposed a context-aware saliency method to detect salient objects. Liu et al. [32] proposed a two-phase graph cut and kernel density estimation approach to detect salient objects. Shen and Wu [33] (Shen) incorporated the concept of low rank matrix to detect the salient objects in the image. Vikram et al. [34] (Vikram) computed local saliency by randomly sampling the image into a number of rectangular regions. Zhang et al. [35] over-segmented the image using mean-shift algorithm and used color compactness feature to yield salient objects. Perazzil et al. [36] opted saliency filters and proposed a model for salient region detection. İmamoğlu et al. [37] (WT) proposed a saliency detection model by extracting low-level features based on wavelet transform. Liu et al. [38] extracted saliency using regional histograms. Yan et al. [39] contributed to the field by proposing a hierarchical saliency detection model specially for handling complex images. Xie et al. [40] (BSLM) proposed a Bayesian saliency technique by utilizing the low and mid level cues. Jiang et al. [41] (AMC) used the concept of absorbing Markov chain to detect salient object present in the image. Singh et al. [42] (SOD-C-PSO) gave a novel approach to linearly combine the different feature maps by estimating the weights using constrained particle swarm optimization. Liu et al. [43] (STREE) proposed a novel saliency tree approach to extract salient objects from the image. Zhu et al. [44] (Zhu) used a multisize superpixel approach based on multivariate normal distribution estimation for salient object detection. Singh and Agrawal [45] (SA) employed a combination of Kullback–Leibler divergence and Manhattan distance to compute the regional feature and an area concept integrated global feature to detect salient objects. Recently Qin et al. [46] (BSCA) introduced the concept of single layer and multiple layer cellular automata to extract salient regions from the image.

The peculiarity of each one of the models is the choice of features, the combination schemes, and many more. As the time passed, researchers started focusing on different aspects of saliency. Some of them believed in enhancing the detection accuracy, while other in improving computation time. It is difficult to handle both simultaneously. In this paper we propose an approach to reduce the computation time to a satisfactory level without degrading the detection accuracy.

Salient regions can be well characterized by the distribution of its local features. The maximal amount of information required to describe a salient object is contained in its corners and edges. The most commonly used descriptor utilizing this idea is the Harris descriptor that uses first gradient as the local feature. Harris descriptor distinguishes among three kinds of regions: flat areas, edges and corners. There are slight signal changes in flat areas, whereas edges and corners contain most of the information. It is observed that the Harris descriptor can determine the corner regions more efficiently than the edge regions. But information regarding edges is equally important. Descriptors using Gabor filters show good performance in capturing edge information. Also a single scale is not always sufficient to yield a local image structure. Keeping this in mind, we used multi-scale Harris and multi-scale Gabor functions to extract corners and edges respectively in the proposed model. From these, handful key points are selected and a convex hull is constructed over it. The convex hull segments the image into two regions: an interior region and an exterior region. The pixels of these regions are used to construct a mixture model, considering the pixels to be drawn from a multivariate kernel function, in order to yield a saliency map.

The proposed model is somewhat similar to the model suggested by Xie et al. [40] where Bayesian saliency is computed by exploiting the low and mid level cues. They extracted courser saliency regions through a convex hull of interest points forming the low level cue. The mid level cue utilizes the concept of superpixels which are further grouped via Laplacian sparse space clustering. The problem with the model suggested by Xie et al. [40] is that it mainly focus on color boosted Harris descriptor to detect keypoints, which may be scattered all over the image, for a coarse region estimation of the salient object. Harris operator may miss few keypoints. The results of Harris descriptor can be seen in Fig. 1(b). The proposed model removes this drawback by utilizing the combination of Harris descriptor and Gabor function. Gabor function is able to determine keypoints which are missed by Harris operator or vice-versa, which can be observed in Fig. 1(d). In addition to it, their model is highly influenced by the clustering results. The prior map computed in Xie et al. [40] model is based on the coarse saliency region and Laplacian sparse subspace clustering method, which is computationally intensive. If the result of the Laplacian sparse subspace clustering is not precise then the prior map may mistakenly include some of the background pixels. Their clustering technique does not always ensure accurate results, as claimed by them in their paper. In the proposed model, the interior region and exterior region of the convex hull are assumed to be drawn from two multivariate Gaussian signals. Expectation maximization (EM) algorithm is used to learn the parameters of the Gaussian signals to determine the saliency map. The EM algorithm is simple and takes less computation time in comparison to clustering method used in Xie et al. [40] model. The proposed model mainly focuses on enhancing the performance of salient object detection in less time.

Experiments are carried out on six publicly available image datasets. The performance of the proposed model and nineteen other state-of-the-art models is evaluated in terms of precision, recall, F-measure, area under curve and computation time.

The paper is organized as follows. Section 2 gives the details of the proposed model. The experimental setup and results are presented in Section 3. Conclusion and future work are included in Section 4.