Saliency of Scene Object

Ashwini Nanaware, Harish Barapatre

Yadavrao Tasgoankoar Institute of Engineering & Technology, Karjat, India
Yadavrao Tasgoankoar Institute of Engineering & Technology, Karjat, India
ashwininanaware19@gmail.com; harish.barapatre@tasgaonkartech.com

Abstract

Computationally detecting salient object from confusing background is most important in computer vision system. Salient object is unique object which stands out different and having closed boundary in space. Saliency at a given location is determined primarily by how different current location is from its surround in color, orientation, motion, etc. There are various approaches of visual attention like top-down approach, bottom-up approach and hybrid approach. Various object detection tasks use attention models very effectively. In this paper, we review three state-of-the-art attention models for their response to scene images. Early results indicate that saliency maps produced by these attention models can be used for salient object detection algorithms.

Keywords – visual attention models, saliency map, state-of-the-art model, salient object detection

1. INTRODUCTION

Human beings can easily identify objects in a natural scene whereas it is always an open challenge for machine vision system. Object detection is requirements for various purposes, such as content based image analysis, image retrieval, object recognition, etc. Before further processing, machine vision system has to extract the salient regions from an unknown background. Extracted saliency maps are widely used in many computer vision applications including object-of-interest image segmentation [4, 5], object recognition [6], adaptive compression of images [7], content-aware image resizing [8, 9, 10], and image retrieval [11].

A lot of research has been done and going on in the vision community on how to extract regions of interest, so called “salient regions” from an image. The majority of existing visual attention approaches is based on the bottom-up computational framework [12, 13] because visual attention is in general unconsciously driven by low-level features in the scene such as intensity, contrast and motion. These approaches consist of the following three steps. The first step is feature extraction, in which multiple low-level visual features, such as intensity, color, orientation, texture and motion are extracted from the image at multiple scales. The second step is saliency computation. The saliency is computed by a center-surround operation [13], self-information [12], or graph-based random walk [2] using multiple features. After normalization and linear/ non-linear combination, a master map [14] or a salient map [15] is computed to represent the saliency of each image pixel.

Top-down approaches of visual attention use prior contextual knowledge of object location and its shape to guide the saliency map. They are task–dependent and based on the fact that search for an object in an image by humans is usually directed and governed by some context, e.g., in the task of searching for pedestrians human will focus their attention on bottom of the image near road rather high up in the sky.

Recently, a hybrid approach of visual attention is projected by various researchers, which attempts to model human attention in a Bayesian framework combining the bottom-up saliency model and top–down contextual information of object location and appearance [16, 17, 18]. This model computes the probability of finding the object at specific location in each image. This probability is estimated by density of filter responses obtained from local image features (bottom-up saliency) for a given image and combine it with the probability density of object shape (shape prior) and object location (location prior) learned from the training samples, in a Bayesian framework.

There are various research methods, which define ‘saliency’. Interest point detectors (IPs) [19, 20] respond to local textured image neighborhood and are widely used for finding image correspondences [20] and recognizing specific object [21]. IPs focuses on individual points. Class specific saliency defines some visual characteristics that distinguish a particular object class (e.g. cars)
from others. Generic saliency measures the saliency of pixels, as the degree of uniqueness of their neighborhood with respect to the entire image or the surrounding area [22, 24]. Saliency maps and visual attention models have been used in many vision tasks such as classification and object detection. In this paper we saw three different methods of visual attention for the task of object detection in natural scenes.

2. MODELS OF VISUAL ATTENTION

A. Harel’s Graph Based Visual Saliency Model

Harel et al. [1] proposed a bottom-up model of visual saliency which uses the same image features as that of Itti’s [13], but defines Markov chains over various image maps and uses the equilibrium distribution over map locations for calculating the activation map (conspicuity map) and saliency maps. In this, natural and efficient saliency computations are achieved by exploiting the computational power, topological structure and parallel nature of graph algorithms. It consists of two steps: first forming activation maps on certain feature channels, and then normalizing them in a way which highlights conspicuity and admits combination with other maps.

They construct a fully connected directed graph joining all the nodes (pixels) of the feature map and assign weight to the edges proportional to the dissimilarity (log ratio of values) between the nodes and their spatial closeness. They define a Markov process on such a graph and estimate the equilibrium distribution of such a chain. The result is an activation map or conspicuity map derived from pair wise contrast. These activation or conspicuity maps are afterward normalized using the same Markovian process, this time constructing the graph from nodes in activation map. The normalized activation maps are later on combined to give a final saliency map.

The human fixation is predicted more reliably by this algorithm than the standard algorithm. Salient regions are detected by this method; even it is far away from object borders. But it gives some problem with multisolution. For better performance this model can be made extensible to multiresolutions.

B. Context aware saliency detection

S. Goferman et al. [2] proposed model of visual saliency which uses four basic principles of human visual attention, which are supported by psychological evidence [24, 25, 26, 27]: local low-level components having contrast and color, global factor, which controls usually occurring features.

The control of the dissimilarity between patches is decreasing, with the increasing of the spatial distance between them. Spatially weighted but also maintains characteristic that different from the norm, visual organization rules states that visual forms may possess one or several centers of gravity about which the form is organized and finally high-level factors, such as human face appearance.

Distinctive colors or patterns are used to obtain high saliency. On the contrary, homogeneous or blurred areas should obtain low saliency values. In this, local-global single scale saliency, they have considered a single patch of scale at each pixel and a pixel is considered salient if the patch centered at pixel is distinctive with respect to all other image patches. Another important thing is positional distance between patches. Patch is salient when the patches similar to it are nearby, and it is less salient when the resembling patches are far away. Dissimilarity measure is proportional to the difference in appearance and inverse proportional to the positional distance. Further multiple scales enhance the saliency. By improving the contrast between salient and non salient regions, multiple scales are decreased the saliency of background pixels. 4 scales: r= {100%, 80%, 50%, 30%} are used and patches are taken.

After that, a large amount of attended localized areas are extracted from the saliency map. Then, each pixel outside the attended areas is weighted according to its Euclidian distance to the closet attended pixel. Finally, the saliency map should be further enhanced using some high level factors, such as recognized objects or face detection algorithm of [28], which generates 1 for face pixels and 0 otherwise.

In image retargeting novel objects and their neighborhood will remain untouched in the resized image. Distortions occur when there is existence of regions of lower significance.

C. Visual saliency detection by spatially weighted dissimilarity

L. Duan et al. [3] proposed a method based on combining two different kinds of information like the dissimilarities between patches which were evaluated in the reduced dimensional space and the spatial distance between patches which was evaluated in the spatial domain. Main objective is to measure the saliency for each patch drawn from an image. Some fundamentals of the saliency are used like dissimilarity, spatial distance and central bias. Candidate salient region is considered when one patch is more different than all the other ones in the reduced dimensional space. Principal component analysis is used to reduce the dimensionality of each image patch which is represented as a vector.

dissimilarity is inversely weighted by the distance. This is done in three core steps i.e. representing image patches, reducing dimensionality and
evaluating the spatially-weighted dissimilarity. Non-overlapping patches from an image are represented as vectors of pixels. Then all patches are mapped into a reduced dimensional space. After that the saliency of each image patch is determined by aggregating the spatially-weighted dissimilarities between patch and all the other ones in the image. The weighting mechanism indicating central bias is use and ultimately, the saliency map is normalized and resized to the scale of the original image, and then is smoothed with a Gaussian filter.

Saliency maps formed by this method are spread out, on the other hand; salient regions are extracted by setting different thresholds for different visual task. The main drawback here is performance degradation which occurs on the multi-scales visual attention regions.

Fig 1: (a) Input image (b) Harel’s GBVS, (c) S Goferman’s saliency map, (d) L Duan’s saliency map
3. CONCLUSION

In this paper, we have reviewed three state-of-the-art models of visual attention for the salient object detection task. The goal of our effort is to see which models of visual attention are best suited for the task of salient object detection in natural scenes. The experimental results showed that S Goferman’s model performed best for salient object detection from complex background. The results clearly show that attention-based models can be used in early stages of salient object detection.

REFERENCES


