J. Vis. Commun. Image R. 43 (2017) 212–226

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

An image texture insensitive method for saliency detection ☆

Avik Hati *, Subhasis Chaudhuri, Rajbabu Velmurugan

Electrical Engineering Department, Indian Institute of Technology Bombay, Mumbai 400076, India

ARTICLE INFO

Article history:
Received 19 July 2016
Revised 31 December 2016
Accepted 2 January 2017
Available online 4 January 2017

Keywords:
Saliency
Texture suppression
Total variation
Sparse segmentation
Relevance feedback
Image matting

ABSTRACT

We propose a texture insensitive, region based image saliency detection algorithm, having excellent detection and localization properties, to obtain salient objects. We use a total variation based regularizer to suppress textures from the image and to make the method invariant to textural variations in the scene. This leads to an image that contains piecewise constant gray valued regions. This texture-free image is sparsely segmented into a small number of regions using the expectation maximization algorithm assuming a Gaussian mixture model. We compute three different saliency measures for every region using its intensity and spatial features. We adopt a relevance feedback mechanism to obtain weights for combining the three saliency measures and obtain the final saliency map. Next we input the thresholded saliency map to an image matting technique and extract the salient objects from the image with exact boundaries. Experimental comparisons with existing saliency detection algorithms demonstrate the superiority of the proposed technique.

1. Introduction

Saliency is a measure of importance of objects or regions in an image (see Fig. 1) or important events in a video scene that capture our attention. The salient regions in an image are different from the rest of the image in certain features (e.g., color or frequency). Salient region detection methods identify important regions in an image so that operations can be performed only on those regions. This reduces complexity in many image and vision applications, which work with large image databases and long video sequences. For example, saliency detection can aid in video summarization, image segmentation, content based image compression, object recognition, image and video quality assessment and progressive image transmission.

The reason this problem is extremely challenging is because the notion of saliency is purely subjective, where we try to address the problem of identifying the elements in an image or scene that capture our attention. Humans prefer to look at an image in a broader sense, e.g., we consider images as a set of objects instead of a set of pixels. Furthermore, we prefer to focus on 'important' objects while giving less attention to the less important ones. The main challenge here is to define the term 'important' in a quantitative sense and subsequently to perform the salient object identification.

A review of some state-of-the-art saliency detection methods is presented in [1,2] and we discuss a few relevant ones here. Frequency domain approaches have been used to detect saliency by exploiting varying spectral components present in the image. The spectral residual (SR) approach [3] combines phase spectrum with the spectral residual of an image to obtain saliency. The phase Fourier transform (PFT) model [4,5] shows that the spectral residual part contains little information about the image. Inverse Fourier transform of only the phase spectrum of the image is taken as the saliency map which highlights the boundaries, but the method gets affected by image textures. PFT also fails if the salient region is large or the background is cluttered. The amplitude information which has been neglected in both SR technique and PFT model, is used to obtain a better saliency map in [6,7]. The amplitude spectrum of the image is lowpass filtered and combined with the phase spectrum to get the saliency map in [6]. The method of Li et al. [7] modulates the phase spectrum using a learned phase filter. Achanta et al. [8] try to retain the salient object boundaries by retaining most of the frequency components in the image. Multiple difference of Gaussians (DoGs) of several narrow passbands are combined to obtain a filter. Saliency of a pixel is computed as the difference between the averaged image and the filtered image. Retention of high frequencies may cause poor result as noise (high frequency) is retained. Ma et al. [9] compute saliency by combining Wavelet transform of different color channels. Fang et al. [10] compute saliency in compressed domain using DCT coefficients. A singular value decomposition based approach has been presented in [11] with the assumption that the large singular values correspond

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

* Corresponding author.
E-mail addresses: avik@ee.iitb.ac.in (A. Hati), sc@ee.iitb.ac.in (S. Chaudhuri), rajbabu@ee.iitb.ac.in (R. Velmurugan).

http://dx.doi.org/10.1016/j.jvcir.2017.01.007
1047-3203/

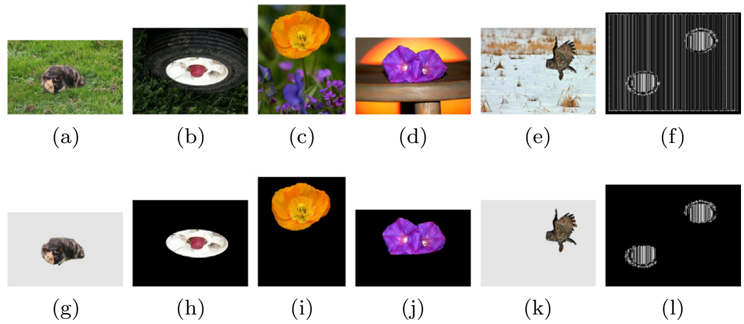

Fig. 1. Example of salient objects. (a)–(f) input images and (g)–(l) the corresponding salient objects obtained using the proposed method. The salient objects in (g) and (k) are shown on gray background and the remaining objects are shown on black background for contrast. (For interpretation of the references to colour in this figure, the reader is referred to the web version of this article.)

to the image background. In spectral methods, the localization of the salient object is quite poor.

Spatial domain approaches based on center-surround color difference by combining color features and edge orientations in different scales of the image have been discussed in [12–16]. The method of Itti et al. [15], one of the early works on saliency, involves computing feature differences between fine and coarse scales of Gaussian image pyramid. A graph based approach [14] tries to solve it using a Markovian approach. Gao et al. [12,13] model saliency of a pixel as the discrimination between feature responses of two windows, a center and a surround window around that pixel, by computing Kullback-Leibler divergence. Similar information theoretic methods have also been used in [17]. But these methods generate only blurred and approximate saliency maps. Kanan et al. [18] and Xie et al. [19] compute saliency based on Bayesian modeling. Learning based methods have been used in [20,21]. Rigas et al. [22], Yan et al. [23], Shen et al. [24], Fareed et al. [25] and Li et al. [26] compute saliency using sparse representation of image. Naqvi et al. [27] use spectral matting for saliency computation. Aytekin et al. [28] and Tian et al. [29] propose a graph-cut based solutions. Whereas Aytekin et al. [28] use concepts from quantum mechanics, Tian et al. [29] incorporate two complementary saliency prior maps in the cost function. These two maps penalize if pixels outside the object are detected as salient and if pixels inside the object are not detected as salient, respectively.

In the visual attention detection technique of Zhai et al. [30], saliency of each pixel is computed as the sum of the distances of rest of the pixels from it. To reduce computation, saliency is computed for each of the 256 gray values by taking weighted sum of distances, where weights are the frequency or histogram values of each gray value. The pixel based local method of Cheng et al. [31] tries to reduce the complexity of the method in [30] when applied to color images by quantizing each of R,G,B channels into 12 bins. The resulting quantization artifacts are smoothed to improve the saliency output to some extent. As an improvement over this, Cheng et al. [31] have also proposed a region based global method for saliency detection that uses image segmentation. Saliency of each image segment is computed as a weighted sum of color distances of remaining regions from it. Perrazi et al. [32] use a similar global method using superpixel segmentation and incorporate terms that consider salient objects to be near the image center, whereas Zhu et al. [33] use foreground and background annotations to incorporate top-down information. Multi-scale saliency detection based on hierarchical segmentation [34] has been proposed in [35–37]. Shi et al. [34] and Xu et al. [35] use color and spatial distances, whereas Zou et al. [36] use surroundedness of a region as a cue for saliency. Jiang et al. [37] learn a regional saliency regressor. Image boundary information has been used as background prior in [28,38–42]. Superpixel based approaches e.g., [43,26,44,45] also solve the localization problem as saliency values are computed for each segment in the image after superpixel segmentation, but the image background is highly over-segmented. As the salient object contains many superpixels that may have different saliency values, the salient object is not uniformly emphasized. This will result in holes inside the salient object and noisy salient pixels outside the salient object after thresholding. This is because some of the superpixels within the object have small saliency values and some superpixels in the background have large saliency values, respectively. Similarly, patch based approaches [46–49] and pixel based methods [8,12–15,30,50–55] also have the above mentioned issues.

In this paper, we propose a novel saliency detection algorithm that uses texture suppression to segment salient objects even in presence of heavy texture. The proposed algorithm (i) aims to detect salient regions irrespective of their textural features, (ii) assigns a *uniform* saliency value to the pixels inside the salient object and extracts the *complete* contour of the salient object unlike pixel based, patch based or superpixel based methods from literature which yield *non-uniform* saliency values, (iii) produces noise-free saliency maps after thresholding without any spurious points around the salient region and without any holes inside it, and does not split a salient object into multiple regions unlike most methods, (iv) unlike all other existing methods [31,32,34,43], eschews all heuristics and derives the weights to combine various types of features from the image itself, (v) focuses on determining whether or not an image contains a salient region at all, and (vi) does not have any salient object localization problem as in spectral approaches.

A very preliminary version of the proposed method has appeared in [56] and the texture suppression part has appeared in [57]. This paper is an enhanced and much expanded version of [56] and further includes additional analyses involving processing of textured images and experimentation on a much larger database.

2. Saliency detection algorithm

The block diagram for the proposed saliency detection algorithm is shown in Fig. 2. The algorithm proceeds as follows. First, textures are suppressed from the image using a total variation (TV) based regularizer to obtain an image with piecewise constant gray valued (PCGV) regions (Section 2.1). With more weightage given to the TV regularizer term, the number of segments obtained in the segmentation step is much less than that obtained without the use of TV or with a less weighted TV regularizer term. The PCGV image is segmented into a very small number of regions by representing it as a mixture of Gaussians (Section 2.3). Then for each region, we compute saliency values using different features (Section 2.4) and use relevance feedback to find appropriate