Hierarchical Saliency Detection

Qiong Yan Li Xu Jianping Shi Jiaya Jia
The Chinese University of Hong Kong
{qyan,xuli,jpshi,leojia}@cse.cuhk.edu.hk
http://www.cse.cuhk.edu.hk/leojia/projects/hsaliency/

Abstract

When dealing with objects with complex structures, saliency detection confronts a critical problem – namely that detection accuracy could be adversely affected if salient foreground or background in an image contains small-scale high-contrast patterns. This issue is common in natural images and forms a fundamental challenge for prior methods. We tackle it from a scale point of view and propose a multi-layer approach to analyze saliency cues. The final saliency map is produced in a hierarchical model. Different from varying patch sizes or downsizing images, our scale-based region handling is by finding saliency values optimally in a tree model. Our approach improves saliency detection on many images that cannot be handled well traditionally.

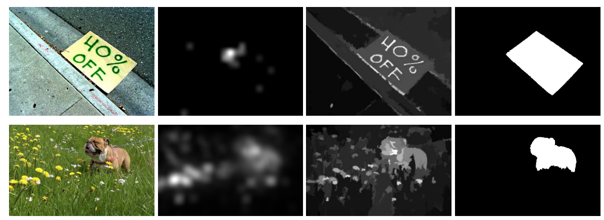
1. Introduction

Saliency detection, which is closely related to selective processing in human visual system [6], aims to locate important regions or objects in images. Knowing where important regions are broadly benefits a set of applications, including classification, retrieval and object co-segmentation, for optimally allocating computation.

By defining saliency as pixel/region uniqueness in either local or global context, existing methods can be classified to two streams. Local methods [3] rely on pixel/region difference in the vicinity, while global methods [1, 2] rely mainly on color uniqueness in terms of global statistics.

Albeit many methods have been proposed, there exist a few commonly noticeable and critically influencing issues, relating to the complexity of patterns in natural images. Two examples are shown in Fig. 1, where either the foreground or background is not uniform, but contains complex structures. Results by a local method [3] and a global method [2] both have difficulties in highlight the foreground uniformly. The issues boil down to a common problem – that is, *when objects contain salient small-scale patterns, saliency could generally be misled by their complexity*.

Aiming to solve this notorious and universal problem,



(a) Input (b) IT [3] (c) RC [2] (d) GT

Figure 1. Saliency detection with structure confusion. Small-scale strong details influence the process and cause erroneous results.

we propose a hierarchical model, to analyze saliency cues from multiple levels of structure, and then integrate them to infer the final saliency map. Our model finds foundation from studies in psychology [5]. It is able to deal with salient small-scale structures, so that salient objects are labeled more uniformly. In addition, contributions in this paper also include a new measure of region scales, which is compatible with human perception on object scales.

2. Hierarchical Model

Our method includes three major steps: image layer extraction, single-layer cue estimation and hierarchical inference as illustrated in Fig. 2. Details are explained below.

2.1. Image Layer Extraction

Image layers, as shown in Fig. 2(b), are coarse representation of the input with different degrees of details, balancing between expression capability and structure complexity. The layer number is fixed to 3 in our experiments. In the bottom level, finest details such as flower are retained, while in the top level large-scale structures are produced.

Starting from an initial over-segmentation, for each layer, we define a threshold and merge segments/regions whose sizes are smaller than the threshold iteratively. The three extracted layers, denoted as $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3$, form a tree structure according to the merging process, as shown in Fig. 2 (e).

To measure region size appropriately with respect to human perception, instead of counting the number of pixels inside a region, we define a new scale measure based on

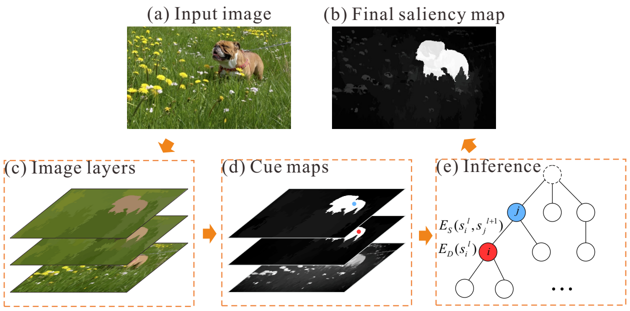
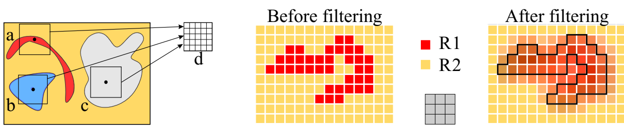


Figure 2. An overview of our hierarchical framework.



(a) (b)

Figure 3. Illustration of region scale. In (a), the scales of regions a and b are less than 5, and that of c is larger than 5. In (b), all values inside region R1 are changed, so its scale is less than 3.

shape uniformities. An illustration is shown in Fig. 3 (a), where the red long curved region with a large number of pixel is not necessarily a large-scale region.

The scale of region R is defined as

$$\text{scale}(R) = \arg\max_{t}\{R_{t\times t}|R_{t\times t} \subseteq R\}, \quad (1)$$

where $R_{t\times t}$ is a $t \times t$ square region, and relation $R' \subseteq R$ means there exists at least one location to put region R' completely inside region R. For efficient computation, we only determine whether the scale of a region is less than t. By applying a box filter with $t \times t$ kernel to the region, if all pixel values inside the region are changed after filtering, it means the region scale is below t, as exemplified in Fig. 3.

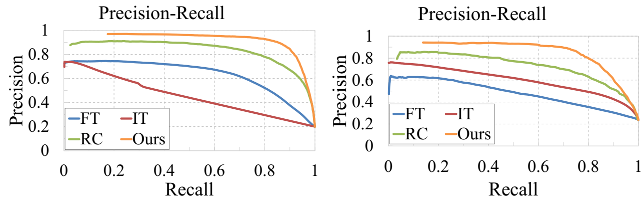
2.2. Single-Layer Saliency Cues

For each layer we extract, saliency cues are applied to find important regions from the perspectives of color, position and size. The saliency cue for a region R_i is determined by local contrast and location heuristic. Local contrast is measured as the sum of color difference from other regions. Location heuristic assign large values to regions close to image center. We denote the calculated value as $\bar{s}_i$.

2.3. Hierarchical Inference

We propose a hierarchical inference procedure to fuse the saliency cue from each layer optimally for multi-scale saliency detection. For a region i in layer $\mathcal{L}^l$, we define a saliency variable s_i^l. Set $\mathcal{S}$ contains all of them. We minimize the following energy function

$$E(\mathcal{S}) = \sum_l \sum_i E_D(s_i^l) + \sum_l \sum_{i, R_i^l \subseteq R_j^{l+1}} E_S(s_i^l, s_j^{l+1}). \quad (2)$$



(a) MSRA-100 (b) CSSD

Figure 4. Quantitative comparison on datasets MSRA1000 and CSSD. Ours is the yellow one in the top.

The energy consists of two parts. Data term, $E_D(s_i^l) = \beta^l||s_i^l - \bar{s}_i^l||_2^2$, is to gather separate saliency cue, where β^l controls the layer confidence. The hierarchical term, $E_S(s_i^l, s_j^{l+1}) = \lambda^l||s_i^l - s_j^{l+1}||_2^2$, is defined for regions pairs satisfying $R_i^l \subseteq R_j^{l+1}$. It makes saliency assignment for corresponding regions in different layers consistent to effectively correcting initial saliency errors. Parameter λ^l controls the strength of consistency.

This model forms a hierarchical tree structure, as shown in Fig. 2 (e), and thus can be efficiently solved via belief propagation with global optimal [4]. After optimizing Eq. (2), we use the optimum in the finest scale, i.e. layer $\mathcal{L}^1$, as our final saliency result. The result in Fig. 2 (b), shows much improvement compared to that in Fig. 1 (b) and (c). More details are available in [7].

3. Experiments

We evaluate our method on the saliency datasets MSRA-1000 [1] and CSSD [7]. Following previous settings in [1, 2], we show quantitative comparison in Fig. 4 with several prior ones, including local method – IT [3], and global methods – FT [1], RC [2]. Our hierarchical method outperforms in both datasets since it integrates saliency information from multiple image scales, obtaining more reliable saliency assignment.

References

[1] R. Achanta, S. S. Hemami, F. J. Estrada, and S. Süsstrunk. Frequency-tuned salient region detection. In *CVPR*, 2009.

[2] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu. Global contrast based salient region detection. In *CVPR*, 2011.

[3] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(11), 1998.

[4] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47(2), 2001.

[5] T. Lee and D. Mumford. Hierarchical bayesian inference in the visual cortex. *JOSA A*, 20(7), 2003.

[6] S. Palmer. *Vision science: Photons to phenomenology*, volume 1. MIT press Cambridge, MA, 1999.

[7] Q. Yan, L. Xu, J. Shi, and J. Jia. Hierarchical saliency detection. In *CVPR*, 2013.