1-2015

Salient Object Detection via Objectness Proposals

Tam Nguyen
University of Dayton, tnguyen1@udayton.edu

Salient Object Detection via Objectness Proposals

Tam V. Nguyen
Department for Technology, Innovation and Enterprise
Singapore Polytechnic, Singapore

Abstract

Salient object detection has gradually become a popular topic in robotics and computer vision research. This paper presents a real-time system that detects salient object by integrating objectness, foreground and compactness measures. Our algorithm consists of four basic steps. First, our method generates the objectness map via object proposals. Based on the objectness map, we estimate the background margin and compute the corresponding foreground map which prefers the foreground objects. From the objectness map and the foreground map, the compactness map is formed to favor the compact objects. We then integrate those cues to form a pixel-accurate saliency map which covers the salient objects and consistently separates fore- and background.

Introduction

The objective of salient object detection is to find the salient objects which draw the attention of humans at the first sight on the image. The recent research has shown that computational models simulating low-level stimuli-driven attention are quite successful and represent useful tools in many practical scenarios, including image retargeting (Avidan and Shamir 2007), image re-attentionizing (Nguyen et al. 2013), and action recognition (Nguyen, Song, and Yan 2015). The existing methods can be classified into biologically-inspired (Itti, Koch, and Niebur 1998) and computationally-oriented approaches (Hou and Zhang 2007; Achanta et al. 2009; Perazzi et al. 2012; Cheng et al. 2013; Bruce and Tsotsos 2005). Despite many recent advancements, the difficult question is still whether "the salient object is a real object". That question bridges the problem of salient object detection into the traditional object detection research. In the object detection, the efficient sliding window object detection while keeping the computational cost feasible is very important. Therefore, there exist numerous objectness proposal generation methods proposing a small number (e.g. 1,000) of proposals, that are expected to cover all objects in an image. Lampert et al. (Lampert, Blaschko, and Hofmann 2008) introduced a branch-and-bound scheme for detection. However, it can only be used to speed up classifiers that users can provide a good bound on highest score. Later, Cheng et al.

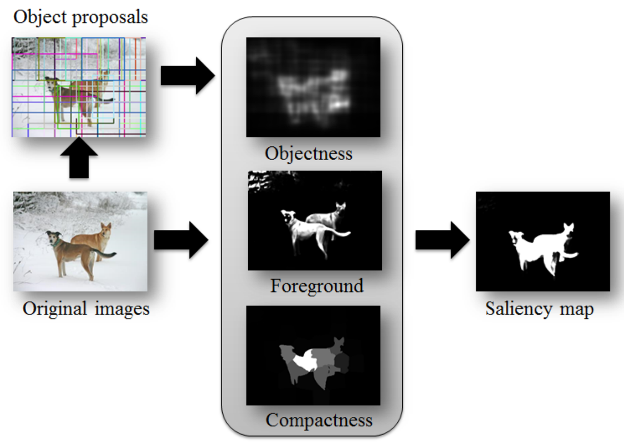

Figure 1: Overview of our system which forms the saliency map from objectness, foreground and compactness maps.

(Cheng et al. 2014) presented a simple and fast objectness measure by using binarized normed gradients features, with which computing the objectness of each image window at any scale and aspect ratio only needs a few bit operations.

The main contributions are three-fold. First, to the best of our knowledge, we integrate objectness proposals into the salient object detection problem. Second, we propose the foreground map and compactness map which can cover both global and local information of the saliency object. Last but not least, our work performs in a real-time manner unlike other works in the literature.

Implementation

In this section, we describe the details of our implementation on how the objectness measures as well as the saliency assignment can be efficiently computed. Figure 1 illustrates the overview of our processing steps. The object proposals are generated from the input image. We apply the state-of-the-art objectness detection technique, e.g., binarized normed gradients (BING) (Cheng et al. 2014), to produce a set of candidate object windows. Our selection of BING is two-fold. First, BING extractor has a weak training from the simple feature, namely, binarized normed gradients. Therefore, it is useful comparing to bottom-up edge

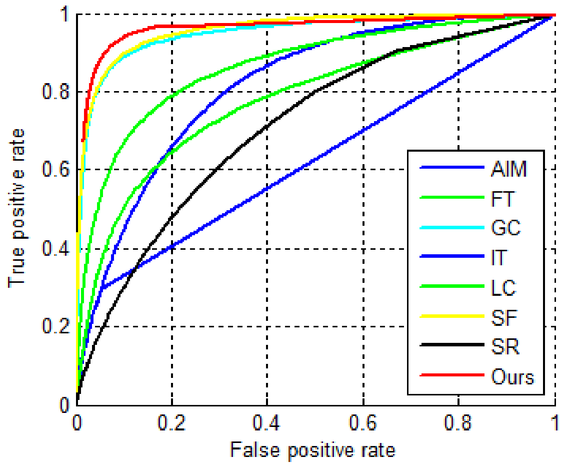

Figure 2: ROC curves of different methods on MSRA-1000 dataset (Achanta et al. 2009).

extractor. Second, the BING extractor is able to run 10 times faster than real-time, i.e., 300 frames per second (fps). BING proposal generator is trained with VOC2007 dataset (Everingham et al. 2010) same as in (Cheng et al. 2014). The objectness map is later formed by accumulating all proposals.

The foreground map is then created from the difference between the pixel's color and the background color obtained following the estimated margins. We then oversegment the image into superpixels and compute the compactness map based on the spatial distribution of superpixels. Regarding the image over-segmentation, we use SLIC (Achanta et al. 2012) for the superpixel segmentation. We set the number of superpixels as 100 as a trade-off between the fine over-segmentation and the processing time. Finally, a saliency value is assigned to each pixel by the multiplication of objectness, foreground and compactness values.

Table 1: Comparison of running times in the benchmark (Achanta et al. 2009).

Method	AIM	FT	SF	GC	Ours
Time (s)	62.4	0.01	0.15	0.09	0.07
Code	MATLAB	C++	C++	C++	C++

Performance Analysis

The widely used ROC curves of our system to other works in MSRA-1000 dataset (Achanta et al. 2009) are shown in Figure 2. Our method outperforms other state-of-the-art baselines (AIM (Bruce and Tsotsos 2005), FT (Achanta et al. 2009), GC (Cheng et al. 2013), IT (Itti, Koch, and Niebur 1998), LC (Zhai and Shah 2006), SF (Perazzi et al. 2012), SR (Hou and Zhang 2007)) In addition, we also compare the average running time of our approach to the currently best performing methods on the benchmark images. The average time is taken by each method on a PC with Intel i7 3.3 GHz CPU and 8GB RAM. As shown in Table 1, our method obtains the fastest running times ($\sim$15 frames per second).

Conclusion

This paper introduces a novel method which adopts the object proposals in order to rapidly detect salient object. Our work considers both foreground and compactness cues, leading to a reliable global and local saliency cues estimation. Our system promises the wider integration of salient object detection into other research works.

Acknowledgments

This work was supported by Singapore Ministry of Education under research Grant MOE2012-TIF-2-G-016.

References

Achanta, R.; Hemami, S. S.; Estrada, F. J.; and Süsstrunk, S. 2009. Frequency-tuned salient region detection. In *CVPR*, 1597–1604.

Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; and Süsstrunk, S. 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* 34(11):2274–2282.

Avidan, S., and Shamir, A. 2007. Seam carving for content-aware image resizing. *ACM Trans. Graph.* 26(3):10.

Bruce, N. D. B., and Tsotsos, J. K. 2005. Saliency based on information maximization. In *Advances in Neural Information Processing Systems 18*.

Cheng, M.; Warrell, J.; Lin, W.; Zheng, S.; Vineet, V.; and Crook, N. 2013. Efficient salient region detection with soft image abstraction. In *ICCV*, 1529–1536.

Cheng, M.; Zhang, Z.; Lin, W.; and Torr, P. H. S. 2014. BING: binarized normed gradients for objectness estimation at 300fps. In *CVPR*, 3286–3293.

Everingham, M.; Gool, L. J. V.; Williams, C. K. I.; Winn, J. M.; and Zisserman, A. 2010. The pascal visual object classes (VOC) challenge. *IJCV* 88(2):303–338.

Hou, X., and Zhang, L. 2007. Saliency detection: A spectral residual approach. In *CVPR*.

Itti, L.; Koch, C.; and Niebur, E. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 20(11):1254–1259.

Lampert, C. H.; Blaschko, M. B.; and Hofmann, T. 2008. Beyond sliding windows: Object localization by efficient subwindow search. In *CVPR*.

Nguyen, T. V.; Ni, B.; Liu, H.; Xia, W.; Luo, J.; Kankanhalli, M. S.; and Yan, S. 2013. Image re-attentionizing. *IEEE Transactions on Multimedia* 15(8):1910–1919.

Nguyen, T.; Song, Z.; and Yan, S. 2015. STAP: Spatial-Temporal Attention-aware pooling for action recognition. *IEEE Trans. on Circuits and Systems for Video Technology*.

Perazzi, F.; Krähenbühl, P.; Pritch, Y.; and Hornung, A. 2012. Saliency filters: Contrast based filtering for salient region detection. In *CVPR*, 733–740.

Zhai, Y., and Shah, M. 2006. Visual attention detection in video sequences using spatiotemporal cues. In *ACM Multimedia*, 815–824.