LETTER

Saliency Density and Edge Response Based Salient Object Detection

Huiyun JING[†], **Qi HAN**[†a)], **Xin HE**[†], *Nonmembers*, *and* **Xiamu NIU**[†], *Member*

SUMMARY We propose a novel threshold-free salient object detection approach which integrates both saliency density and edge response. The salient object with a well-defined boundary can be automatically detected by our approach. Saliency density and edge response maximization is used as the quality function to direct the salient object discovery. The global optimal window containing a salient object is efficiently located through the proposed saliency density and edge response based branch-and-bound search. To extract the salient object with a well-defined boundary, the GrabCut method is applied, initialized by the located window. Experimental results show that our approach outperforms the methods only using saliency or edge response and achieves a comparable performance with the best state-of-the-art method, while being without any threshold or multiple iterations of GrabCut.
key words: *salient object detection, maximum saliency density and edge response, branch-and-bound search*

1. Introduction

The human visual system always automatically attends to salient objects. This ability enables us to allocate limited processing resources on important parts of an image. Salient object detection has many applications in machine vision [1], [2]. Recently, many studies have been published concerning it, and these can be divided into two categories.

The first category intends to find the most possible rectangle window containing a salient object. Liu et al. [3] used exhaustive search to find the optimal window from the thresholded saliency map. To avoid the brute force search, Valenti et al. [4] applied efficient sub-window search (ESS) [5] to speed up the search process. Nevertheless, the results of the both methods [3], [4] highly rely on the threshold choice. To get rid of threshold, Luo et al. [6] used the maximum saliency density as the quality function of ESS method to locate the salient object. Shi et al. [7] iteratively performed ESS with an objective function of region diversity maximization (RDM) on the raw saliency map to detect the salient object. This achieves more precise results.

The second category employs object segmentation to obtain the salient object with a well-defined boundary. Achanta et al. [8] selected the regions with higher saliency than an adaptive threshold as the salient object. For more accurate results, Cheng et al. [9] employed the GrabCut [10] method to segment salient object, initialized with the binary saliency map using a fixed threshold. However, owing to the not very appropriate input for initializing GrabCut, multiple iterations of GrabCut and extra dilation and erosion operations to update the input for each iteration are necessary for Cheng's method [9] to help an accurate segmentation.

The above review of the literature reveals that several specific issues remain to be addressed. First, only saliency information is used to guide the salient object location. However, much information in the original image is inevitably lost during the saliency computation, which may include the important cues providing support for salient object detection. Second, to refine the salient object extraction results, multiple iterations of GrabCut and extra operations (dilation and erosion) are necessarily performed in [9], which sacrifices efficiency. Third, a threshold is indispensable in these methods [3], [4], [9] to generate the binary saliency map, from which the salient object is further detected. Hence the performance of those methods deeply depends on the threshold choice and it is difficult to select an appropriate threshold value.

To address the above mentioned issues, we propose a novel threshold-free salient object detection method which simultaneously uses saliency and edge response to guide the salient object detection. In the proposed approach, the salient object is located by searching an optimal window of maximum saliency density and edge response on the raw saliency map rather than the binary saliency map. Then, for further extracting the salient object with a well-defined boundary, the GrabCut method is applied, after being initialized by the located window, without any threshold, multiple iterations of GrabCut, or extra operations to update the input for each iteration.

In this letter, the limitations of existing methods are described in Sect. 2. Section 3 introduces the proposed salient object detection approach. Sections 4 and 5 are respectively devoted to experimental results and conclusion.

2. Limitations of Existing Methods

Salient objects are regarded as the foreground objects which attract more visual attention. Saliency, representing visual attention, is an important cue to direct the salient object detection. Existing methods [3], [4], [6], [7] only using saliency have achieved promising results, especially RDM [7]. However, due to the insufficiency of saliency models, salient object regions are not completely highlighted or background regions are not sufficiently suppressed in the saliency maps [7]. It leads to saliency

Manuscript received October 19, 2012.
Manuscript revised January 1, 2013.
[†]The authors are with the Department of Computer Science and Technology, Harbin Institute of Technology, Harbin, China.
a) E-mail: hanqi_xf@hit.edu.cn
DOI: 10.1587/transinf.E96.D.1243

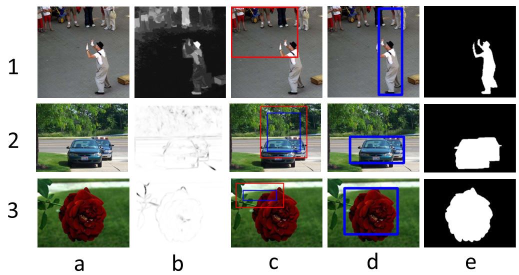

Fig. 1 (1) failure case of method [7], (2) and (3) challenging examples for the method [11], (a) input image, (1-b) saliency map, (2-b) and (3-b) edge response map, (1-c) window search result using [7], (2-c) and (3-c) window search result using [11], (d) window search result using our search algorithm, (e) ground truth.

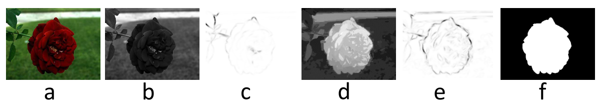

Fig. 2 Comparison of edge response maps generated using MER [11] and our approach (ERSM). (a) input image, (b) grayscale intensity image, (c) edge response map generated by MER [11], (d) saliency map [12], (e) edge response map generated by our approach (ERSM), (f) ground truth.

misguiding the salient object detection (Fig. 1 (1)).

Gu et al. [11] observed that objects are always surrounded by more elongated edges than they have in their interior. Based on the observation, they addressed the object detection problem by searching for the frame-like region that contains the most edge responses (MER). However, only using edge response to detect salient object, MER can't effectively handle the case where background regions have strong edge responses (Fig. 1 (2)). In addition, edge responses are computed from the grayscale intensity image of the truecolor image in MER. It tends to weaken the edge responses around the boundary of salient object, when the intensity values of salient object and image background are similar (Fig. 1 (3) and Fig. 2 (b)).

To solve the above-mentioned problems, saliency and edge response are simultaneously taken into consideration to perform window search in our approach, and the saliency map generated by our previous work [12] is considered as the input to compute edge responses. The salient objects are highlighted from background on our saliency maps, which helps to get stronger edge responses around the boundary of salient object.

3. Salient Object Detection

Different from the aforementioned methods, a novel threshold-free salient object detection approach based on saliency density and edge response is presented in this Section. In the proposed approach, a new quality function integrating saliency density and edge response is devised to guide the salient object detection. Based on the quality function, the optimal window containing a salient object is located by performing branch-and-bound search. Then the located window is used to initialize GrabCut and the salient object with a well-defined boundary is extracted after running GrabCut.

3.1 Maximum Saliency Density and Edge Response (MSDER)

1) Problem Formulation: We model salient objects in images as having higher saliency density within their interior than other regions, while being surrounded by more elongated edges than they have in their interior. The salient object is detected by searching a rectangle window, which simultaneously satisfies the two requirements.

Let I be the image, $W \subset I$ be the searching window, $W' \subset W$ be the inner rectangle of W (the sides of W' live at a certain distance to W) and S be the saliency map. The optimal window $\hat{W}$ is searched with

$$\hat{W} = \arg\max_{W \subset I} F(W) \tag{1}$$

$$F(W) = F_e(W) * F_d(W) \tag{2}$$

$$F_e(W) = \sum_{p \in W} f(p) - \sum_{p \in W'} f(p) \tag{3}$$

$$F_d(W) = \frac{\sum_{p \in W'} S(p)}{A(W')} \tag{4}$$

where $F_e(W)$ and $F_d(W)$ respectively represent the edge response scores computed with Eq. (3) and the saliency density computed by Eq. (4), p denotes a pixel, $f(p)$ expresses edge response for each pixel, $A(W')$ represents the area of W'. Inspired by Gu et al. [11], we also use the frame-like region bounded by two nested windows $\{W, W'\}$ to represent the region occupied by the salient object boundary. So the edge response scores are computed within the border region between W and W' by Eq. (3). The inner window is regarded as the interior region of salient object. Thus the saliency density is computed from W'.

2) Edge Response of Saliency Map (ERSM): We use the saliency map generated by our previous work [12] as the input to compute edge response. In [12], global color contrast and spatial distribution of color are used as the saliency measures. Under this scheme, salient objects are effectively highlighted from image background on the saliency maps (Fig. 2 (d)). Thus the intensity values of salient object and background in the saliency maps [12] tend to be significantly different, which is beneficial to enhance the edge responses of salient object (Fig. 2 (e)).

Given saliency map S and a bank of 15 Gabor filters $\{G_i\}_{i=1}^{15}$, then the edge response for each pixel $f(p)$ is computed with:

$$f(p) = \max_{i=1}^{15} |(S * G_i)(p)|^2 \tag{5}$$

where $*$ is the convolution operator and $|\cdot|$ is the norm of a complex number. The edge responses obtained from saliency maps [12] better represent the boundary of salient object than the ones of grayscale intensity images (Fig. 2).

3.2 Window Search Algorithm

Recently, ESS [5] based on branch-and-bound search scheme is used to replace the exhaustive search for speeding up the salient object detection process. Though ESS is a general search tool, the formulated objective function for optimization is the key to affect the search accuracy. A novel quality function different from that in original ESS method is proposed in our approach. Thus to execute the subsequent branch-and-bound search process, we need to construct the upper bound of our quality function $F(W)$.

For a set of rectangles $\mathbb{W}$, we denote by W_{max} the largest rectangle and by W_{min} the smallest in $\mathbb{W}$. The upper bound $\hat{F}(\mathbb{W})$ of Eq. (1) can be estimated as:

$$\hat{F}(\mathbb{W}) = \left[\sum_{p\in W_{max}} f(p) - \sum_{p\in W'_{min}} f(p)\right] * \frac{\sum_{p\in W'_{max}} S(p)}{A(W'_{min})} \qquad (6)$$

where W'_{max} and W'_{min} respectively are the inner windows of W_{max} and W_{min}. A salient object always accounts for a significant portion of the image. We assume that salient object covers one quarter of the whole image, and the side of object bounding window is one half of the image side. So the distance between the searching window and its inner window is set to one eighth of the image maximum side in our approach. The upper bound $\hat{F}(\mathbb{W})$ has the desired properties [5]: i) $\hat{F}(\mathbb{W}) \geq F(W), \forall W \in \mathbb{W}$ and ii) $\hat{F}(\mathbb{W}) = F(W)$, if W is the only element in $\mathbb{W}$. It guarantees our window search algorithm, according to the search scheme of ESS, converges to a globally optimal solution.

3.3 Salient Object Extraction

After executing the above window search, the optimal window containing a salient object is located. However, the window can not specify the exact shape of salient object. To get the salient object with a well-defined boundary, the GrabCut method is applied in our approach.

We initialize GrabCut using the located window, rather than the binary saliency map generated by a threshold. More specifically, the region inside the located window is set to possible foreground and the remaining areas are set to background. After initializing, GrabCut is performed (one iteration in our approach) to extract the salient object with a well-defined boundary, without multiple iterations of GrabCut or extra operations to update the input for each iteration.

4. Experimental Results

We perform experiments on a publicly available dataset provided by Achanta et al. [8]. The dataset provides 1000 images and associated ground truth in the form of accurate human-labeled masks for salient objects. We compare our approach (MSDER) with four state-of-the-art salient object detection methods, including RDM [7], FTMS [8], RCC [9] and MER [11]. Average precision, recall, and F_α ($\alpha = 0.5$) are used as the performance measures.

1) Comparison of MSDER with RDM: First of all, we compare MSDER with RDM [7]. Saliency map has great impact on the accuracy of salient object detection [6], [7]. RDM gets better performance on Hou's saliency map (SR) [13] than on Achanta's [8] and Bruce's [14] saliency maps. To make a fair comparison, we evaluate RDM on both Hou's (SR) and our saliency maps (DR) [12], and apply our proposed salient object extraction method to detect the salient object with a well-defined boundary. Due to good performance of DR, RDM (DR) obtains more accurate results than RDM (SR) as shown in Fig. 3 and Fig. 6. Furthermore, our approach MSDER, simultaneously taking into account the saliency density and edge response, outperforms both RDM (SR) and RDM (DR) (see Fig. 3 and Fig. 6).

2) Comparison of MSDER with MER: We compare MSDER with MER [11]. Since background regions have stronger edge responses than salient objects (Fig. 4 b), MER does not precisely detect the salient object (Fig. 4 c and d). Compared with MER, our results are more consistent with ground-truth (Fig. 4 g). It is further validated by the precision, recall and F_α results on Fig. 6.

3) Comparison of MSDER with FTMS and RCC: FTMS [8] and RCC [9] share the same purpose with our approach, detecting salient object with a well-defined boundary. For comprehensive comparison, our method is not only compared with RCC (multiple iterations of GrabCut are performed), but also with $RCC_{oneiter}$ (only one iteration of GrabCut is performed). As we can see in Fig. 5 and Fig. 6, our approach (MSDER) performs better than FTMS and $RCC_{oneiter}$. And our approach yields a comparable performance with RCC, without any threshold, multiple iterations

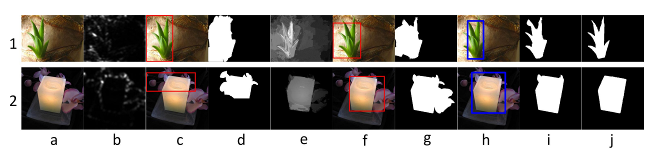

Fig. 3 Comparison of salient object detection results generated using RDM [7] and our approach (MSDER). (a) input image, (b) Hou's saliency map (SR), (c) window detection result (executing RDM on (b)), (d) object detection result (extracting salient object based on (d)), (e) our saliency map (DR), (f) window detection result (executing RDM on (e)), (g) object detection result (executing salient object extraction on (f)), (h) window detection result using our approach (MSDER), (i) object detection result of our approach (MSDER), (j) ground truth.

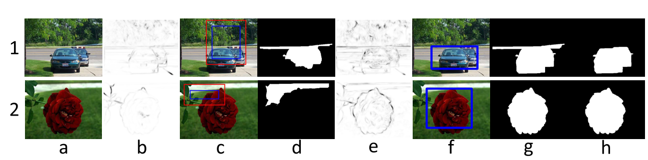

Fig. 4 Comparison of salient object detection results generated using MER [11] and our approach (MSDER). (a) input image, (b) edge response map generated by MER [11], (c) window detection result of MER, (d) object extraction result of MER, (e) edge response map generated by our approach (MSDER), (f) window detection result of our approach (MSDER), (g) object extraction result of our approach (MSDER), (h) ground truth.

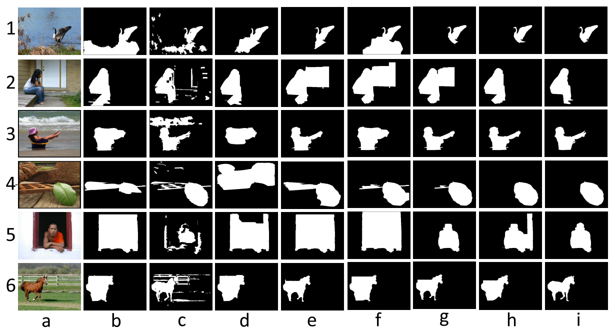

Fig. 5 Comparison of salient object detection results generated using RDM [7], FTMS [8], MER [11], RCC [9] and our approach (MSDER). (a) input image, (b) detection result executing RDM on SR saliency map, (c) detection result of FTMS, (d) detection result of MER, (e) detection result of $RCC_{oneiter}$, (f) detection result executing RDM on DR saliency map, (g) detection result of RCC, (h) detection result of our approach (MSDER), (i) ground truth.

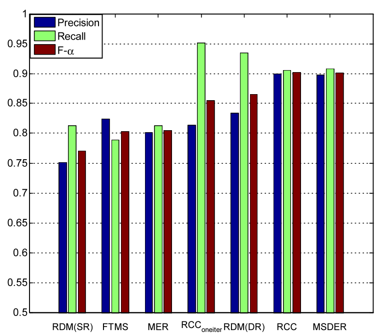

Fig. 6 Quantitative comparison of different methods on the dataset [8].

Table 1 Average and standard deviation of processing time.

Method	**MSDER**	$RCC_{oneiter}$	**RCC [9]**
Average Time (s)	1.355	1.106	2.877
Standard deviation	0.013	0.011	0.016
Code	C++	C++	C++

of GrabCut, or extra operations to update the input for each iteration. Furthermore, we compare the average and standard deviation of processing times taken by our method (MSDER), $RCC_{oneiter}$ and RCC to detect salient object for images in the dataset [8] (see Table 1). The average processing time of our method is near to $RCC_{oneiter}$, and only one half of that needed by RCC. Since the proposed window search algorithm is carried out in our approach, our approach is slightly slower than $RCC_{oneiter}$.

5. Conclusions

A novel threshold-free approach is proposed to detect the salient object with a well-defined boundary. We formulate the problem as an optimization problem to find a window with maximum saliency density and edge response scores. Then the global optimal window is located by applyingthe proposed saliency density and edge response based branch-and-bound search. After initializing GrabCut with the located window, it is performed to detect the salient object with a well-defined boundary. The experiments demonstrate that the proposed approach outperforms them only using saliency or edge response. Meanwhile, it achieves a comparable performance with the best state-of-the-art method, without any threshold, multiple iterations of GrabCut, or extra operations.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (61100187) and the China Postdoctoral Science Foundation (2011M500666).

References

[1] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition?," 17th IEEE Conference on Computer Vision and Pattern Recognition, pp.37–44, 2004.
[2] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.2376–2383, 2010.
[3] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Shum, "Learning to detect a salient object," IEEE Trans. Pattern Anal. Mach. Intell., vol.33, no.2, pp.353–367, 2011.
[4] R. Valenti, N. Sebe, and T. Gevers, "Image saliency by isocentric curvedness and color," IEEE International Conference on Computer Vision, pp.2185–2192, 2009.
[5] C. Lampert, M. Blaschko, and T. Hofmann, "Efficient subwindow search: A branch and bound framework for object localization," IEEE Trans. Pattern Anal. Mach. Intell., vol.31, no.12, pp.2129–2142, 2009.
[6] Y. Luo, J. Yuan, P. Xue, and Q. Tian, "Saliency density maximization for efficient visual objects discovery," IEEE Trans. Circuits Syst. Video Technol., vol.21, no.12, pp.1822–1834, 2011.
[7] R. Shi, Z. Liu, H. Du, X. Zhang, and L. Shen, "Region diversity maximization for salient object detection," IEEE Signal Process. Lett., vol.19, no.4, pp.215–218, 2012.
[8] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," 22nd IEEE Conference on Computer Vision and Pattern Recognition, pp.1597–1604, 2009.
[9] M. Cheng, G. Zhang, N. Mitra, X. Huang, and S. Hu, "Global contrast based salient region detection," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.409–416, 2011.
[10] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," ACM Trans. Graphics, vol.23, no.3, pp.309–314, 2004.
[11] S. Gu, Y. Zheng, and C. Tomasi, "Saliency and localization by nested window search," Tech. Rep. CS-2011-07, Duke University, June 2011.
[12] X. He, H. Jing, Q. Han, and X. Niu, "Salient region detection combining spatial distribution and global contrast," Optical Engineering, vol.51, no.4, pp.047007-1–047007-4, April 2012.
[13] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," IEEE Conference on Computer Vision and Pattern Recognition, pp.1–8, 2007.
[14] N. Bruce and J. Tsotsos, "Saliency, attention, and visual search: An information theoretic approach," J. Vision, vol.9, no.3, p.5, 2009.