SALIENCY DETECTION BASED ON EXTENDED BOUNDARY PRIOR WITH FOCI OF ATTENTION

Yijun Li[1], *Keren Fu*[1], *Lei Zhou*[1], *Yu Qiao*[1], *Jie Yang*[1*] , *and Bai Li*[2]

[1]Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, China
[2]University of Nottingham, UK

ABSTRACT

In this paper, we propose a novel bottom-up paradigm for detecting visual saliency. Regarding the boundary as potential background (boundary prior), we firstly transfer the input color image into a graph with additional four virtual nodes. With a new type of edge called *feature edge* defined considering both color information and spatial distribution, geodesic saliency measure is used to obtain four saliency maps. Then a combination strategy of four maps is proposed, rendering a uniform saliency map to better suppress background and avoid over-suppression of salient object. Finally, we introduce a way of determining foci of attention based on *maximal deviation from norm* (MDN) to enhance the quality of saliency map. Experimental results on a benchmark dataset demonstrate the better performance of our proposed approach compared with several state-of-art methods.

Index Terms— Saliency detection, Saliency map, Boundary prior, Combination of maps, Foci of attention

1. INTRODUCTION

Generally, a salient object is defined as something that captures human perceptual attention. With the help of the human visual system, people can identify salient objects easily from a scene that is even very complex. This magic visual attention mechanism arouses a lot of research interest and much work has been done to realize this ability in computer vision systems. In general, saliency detection approaches usually can be categorized into two groups, i.e., bottom-up and top-down. Bottom-up category [1, 2, 3, 4, 5, 6, 7, 8, 9] simulates our instinctive visual attention mechanism and employs lots of low-level features like color (intensity) and edge (texture) without any scene understanding. Top down category [10] actually requires more prior knowledge and high-level information like face and text. Our approach belongs to the former one.

Bottom-up category methods are data-driven and based on some priors. The most fundamental prior is color contrast [1, 2, 4, 6], which means that the color components belonging to a salient object often have strong contrast to their surroundings. However, such a hypothesis usually results in attenuated inner part for large scale objects and it is also limited in suppressing background clutter. Therefore, other priors are proposed to help obtain better detection results such as center prior [6, 9, 11], rarity prior [12] and shape prior [11].

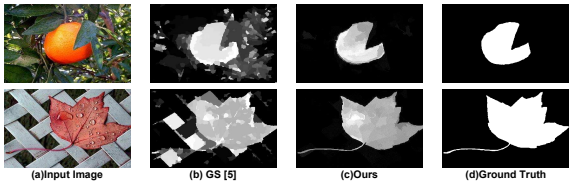

Fig. 1: Existing problems. From left to right: Input image, GS [5], Ours and Ground Truth.

Boundary prior is recently proposed by [5], pointing to the fact that the image boundary is mostly background since photographers usually will not crop salient objects along the view frame. The significance of this prior shifts our focus from what is the salient object to what is the background. In their method, the initial non-salient background parts are propagated to inner parts along a path with the shortest geodesic distance. The larger the distance is, the more salient the part will be. However, it has one main limitation on background suppression due to the effect of accumulation along the path.

To be specific, since the edge set of their graph only comprises of edges connecting geographically neighboring pixels, non-salient inner parts will get relatively high saliency. As illustrated in the 1st row of Fig. 1(b), the cluttered leaves around the orange are not suppressed as well as those near the boundary. More seriously, when some non-salient inner parts has high color contrast against each other along the path, things may get even worse as shown in the 2nd row of Fig. 1(b). The dark green parts between fences are assigned even higher saliency than the salient leaf itself.

Motivated by the aforementioned problems and considerations, we propose a method that could suppress background clutter as well as pop out salient object uniformly. The main contributions of our approach lie in three aspects:

(i) The graph is constructed with additional newly-defined nodes and edges to derive four saliency maps.

(ii) A combination strategy is put forward to better sup-

*Corresponding author: Jie Yang, jieyang@sjtu.edu.cn

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

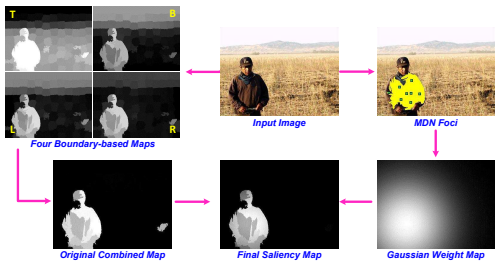

Fig. 2: The framework of our approach

press background and avoid over-suppression of salient object, contributing to a uniform saliency map.

(iii) A way of determining foci of attention based on MDN is introduced to further enhance the quality of saliency map.

The rest of this paper is organized as follows. Our approach is described in Section 2. Experimental results and comparisons are showed and analyzed in Section 3 and conclusions are given in Section 4.

2. THE PROPOSED METHOD

The framework of our method is shown in Fig. 2. Since pixels in the same region usually have homogenous color component, we choose superpixel [13] rather than pixel as the smallest processing unit to decompose an image and generate K spatial compact regions first. Each superpixel was labeled as i ($i = 1, 2, \ldots K$) and its mean position and mean color in LAB color space are denoted as p_i and c_i respectively (both normalized to the range [0,1]).

2.1. Graph Construction

We build an undirected weighted graph G=(V, E), where $V = \{I \cup \Phi\}$ is the node set and E is the edge set. Each superpixel corresponds to a node in I and four virtual nodes are additionally incorporated, denoted as $\phi_l \in \Phi$ where $l \in (top, bottom, left, right)$. The edge set $E = \{(i, j)\}$ includes three types of edge: *adjacent edge*, *virtual edge* and *feature edge*. Fig. 3 presents an example of our graph.

Adjacent Edge: The *adjacent edge* denotes the connection of pairs of superpixels that are neighbors since adjacent superpixels have greater chances to be on a similar saliency level than those apart. The weight is defined by the Euclidean distance of two nodes in LAB color space.

Virtual Edge: The *virtual edge* targets on dealing with the situation when salient object tends to be partly cropped on the boundary. When this happens, it is observed that the boundary superpixels on the object are more salient than boundary superpixels in the background. Therefore, the weights of *virtual edge* computation is treated as a one-dimensional saliency detection problem and we adopt the same measure in [5]. Here our main contribution is to employ four virtual nodes

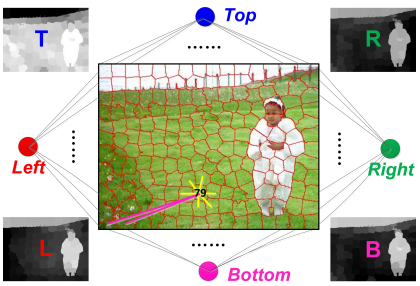

Fig. 3: Our graph model: for node (superpixel) $No.79$, the yellow, pink and black line is the *adjacent edge*, *feature edge* and *virtual edge* respectively.

rather than one and connect ϕ_l with superipixels on l boundary where $l \in (top, bottom, left, right)$. The reason why we do this is to prepare for deriving four saliency maps (discussed in Section 2.2) and with the combination strategy, the goal of background suppression will be realized.

Feature Edge: The *feature edge* is introduced to effectively connect feature-similar parts which are not adjacent. Therefore it can help reduce the geodesic distance because once a background part finds out a nearer neighbor in feature space, it can stride over dissimilar parts surrounding it. In order to estimate the *feature edge* of a node, its N-nearest neighboring nodes in LAB color space are evaluated. However, for a salient object, its color components usually distribute compactly and similar superpixels are less likely to be of one object if they are spatially far away. Therefore we also take the spatial information into consideration and define *feature edge* in the following steps:

(i). For each superpixel i, find the nearest superpixel t in LAB color space.

(ii). The weight of *feature edge* between i and t is defined as:

$$\omega(i, t) = e^{\beta \|p_i - p_t\|_2} \times \|c_i - c_t\|_2 \quad (1)$$

which indicates that if two similar parts are too faraway, though we connect them, the larger weight will make such an edge less likely to contribute to the final shortest path (discussed in Section 2.2).

(iii). Find the next nearest superpixels $t^{'}$ and repeat (ii) until we finish the N-nearest evaluations.

2.2. Combination of Multiple Saliency Maps

With the graph, the geodesic saliency of a superpixel i is measured by the summation of edge weights along the shortest path from i to the virtual node. Taking top boundary for example, for each node i we compute its top saliency S_1 by the following process:

$$S_1(i) = \min_{N_1=i, N_2, \ldots, N_{end}=\phi_{top}} \sum_{k=1}^{n-1} \omega(N_k, N_{k+1}), \quad s.t.(N_k, N_{k+1}) \in E \quad (2)$$

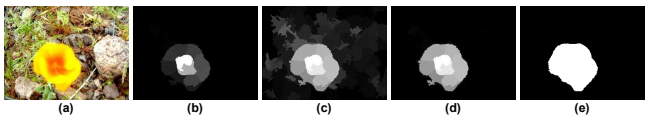

Fig. 4: The saliency map comparisons for different combination strategies of four saliency maps. From left to right: Input image, Simply multiplication, Simply the average, Ours and Ground Truth.

where the shortest path is computed by *Dijkstra's* algorithm. The same process goes for the bottom, left and right boundary and thereby rendering four saliency maps as shown in Fig .3, normalized to [0,1].

Then we set a threshold $\hat{S_h}$ by employing OTSU algorithm [14] as a criterion to distinguish salient object and background in each map. Lastly a combination strategy of four maps is proposed to obtain the final saliency map $S_{boundary}$ as follows:

$$S_{boundary}(i) = \begin{cases} \frac{1}{4}\sum_{h=1}^{4} S_h(i), & if\ \forall\ h\ \ S_h(i) > \hat{S_h} \\ \prod_{h=1}^{4} S_h(i), & otherwise \end{cases} \quad (3)$$

where S_1, S_2, S_3 and S_4 represent the top, bottom, left and right boundary saliency map respectively.

The multiplication strategy plays the main role of background suppression. Since all values in maps are normalized to range [0,1], the multiplication of four values will be much less than any of the four. It works well as what we hope for background saliency, i.e., the smaller, the better. However in such a way of simply multiplication, salient parts will also be affected, namely being over-suppressed. Fig. 4(b) clearly shows what we concern about. To fully highlight the flower in Fig. 4(a), we adopt the average strategy to avoid the over-suppression of salient object. Iff the saliency value of a superpixel is higher than the corresponding $\hat{S_h}$ in four maps, it will be regarded as to be salient and its saliency is obtained by averaging the corresponding four values. As shown in Fig. 4(d), our strategy renders a more satisfactory result.

2.3. Foci of Attention: Maximal Deviation from Norm

To further enhance our results, we next incorporate a known prior on image organization. According to Gestalt laws [15], visual forms may possess one or several centers of gravity about which the form is organized. This suggests that areas that are close to the foci of attention should be cared more than faraway regions. Usually, the image center is chosen as the focus of attention [6, 9, 11]. Obviously this is not always true because salient object can be placed anywhere in image scenes. Here we propose a novel way to determine some locations in an image as the foci of attention and employ them to enhance the quality of saliency map.

Algorithm 1 MDN foci of attention algorithm.

Input: Mean color c_i of each superpixel i of pre-segmented color image A, the number of superpixels K and the number of foci M;
Output: Foci of attention set Ω;

```
Initial Ω = {};
repeat
    Compute the average color C_m of all c_i in A\Ω;
    Compute the Euclidean distance dis(i) between c_i and
    C_m in LAB color space;
    Find the superpixel T satisfying dis(T) = max(dis);
    Add T in Ω, M = M − 1;
until M = 0;
```

The fundamental principle behind our algorithm is that the saliency of a pixel depends on how largely it deviates from the norm (average) of the image in LAB color space. The larger the deviation is, the more likely it is to be salient. We adopt an iterative way to obtain M foci of attention. In each iteration, the superpixel that has the maximal deviation from norm (MDN) is popped out and is eliminated from image before next iteration. Algorithm 1 shows this proposal in details.

2.4. Final Saliency Assignment

With MDN foci of attention, we place a Gaussian with $\sigma = 100$ at the center of each superpixel $i \in \Omega$ and generate a corresponding weight map G_i. Our final saliency map S is defined as follows:

$$S(i) = S_{boundary}(i) \times G(i) \quad (4)$$

where $G = \frac{1}{M}\sum_{j=1}^{M} G_j$. Fig. 2 presents an example of our Gaussian weight map. Every popped-out superpixel is painted in yellow and its mean position p_k is drawn as a green square. Here we eliminate a few foci that are too close to the boundary of the image. It is observed that MDN foci points of attention lie on the salient man and the saliency map is enhanced by suppressing the artifact on the right side. This is in accord with our prior because the artifact is faraway from foci of attention.

3. EXPERIMENT AND COMPARISON

3.1. Performance Comparisons

We test our method on a popular benchmark dataset which includes 1000 images and their manually labeled ground truth [3]. In the following experiments, each image is pre-segmented into about $K = 100$ superpixels. Other three parameters are all empirically chosen as: $\beta = 5$, $N = 4$ nearest neighboring nodes for *feature edge* and $M = 10$.

We evaluate the results of our method against six state-of-art bottom-up saliency detection methods: the IT [1], FT

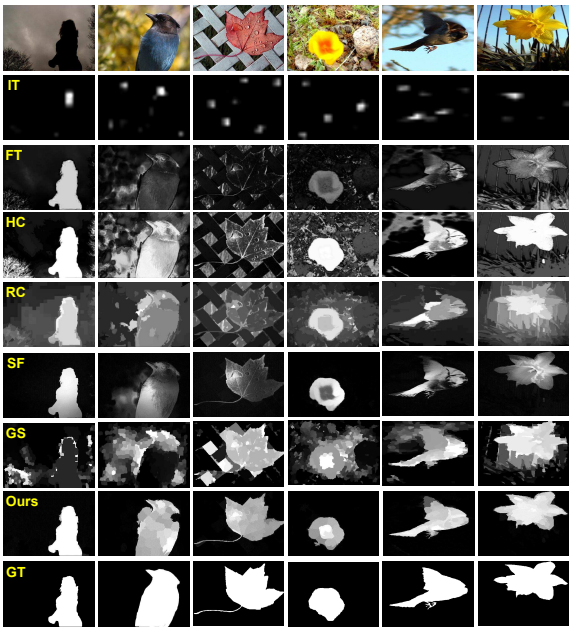

Fig. 5: Visual performance comparisons of the proposed method and the six existing methods. From top to down: Input image, IT, FT, HC, RC, SF, GS, Ours and Ground Truth.

[3], HC [2], RC [2], SF [4] and GS [5]. Visually, Fig. 5 demonstrates the validity and efficiency of our method. It can be seen that our method gains higher quality saliency maps compared with other methods. The 1st and 2nd column show that our method also works well in cases of the salient object cropped on boundary. The following four columns verify our superiority on suppressing multiple kinds of background. Especially for the 3rd column which is a failure case mentioned in [5], we detect the red leaf nearly as good as the ground truth.

Similar as [2, 4, 6], we also implement quantitative evaluation by employing two criteria: one is to compute the averaged precision-recall curves by binarizing the saliency map using thresholds ranging from 0 to 255; the other is to apply F-measure by integrating both precision and recall. Fig. 6 presents the precision-recall curves and F-measure curves of all methods.

From Fig. 6(a), it is noted that the proposed method reaches the best. Especially compared with GS [5] (the red dotted line), our maximum precision rate is 96% with nearly 6% improvement relatively. Due to the background suppression, when the racall is high (corresponding to low threshold), our method still reaches the highest precision because the saliency value of background is still under the threshold. Besides, Fig. 6(b) shows that our approach achieves the best for a very large range $T \in [0, 200]$ with the highest F-measure score 0.89. For $T > 200$, GS and HC method are a bit higher because of their relatively higher recall under a large threshold, which suggests that in the aspect of highlighting salienct object, they show a little advantage. As mentioned before, our approach focuses more on background suppression.

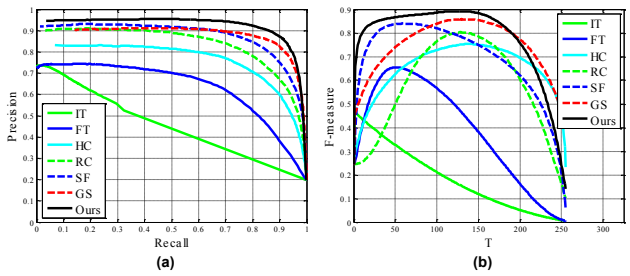

Fig. 6: Quantitative comparisons for the proposed method and six existing method: (a) Precision-Recall curves; (b) F-measure curves.

Methods	RC [2]	SF [4]	Ours	GS [5]	CA [9]
Time(s)	0.254	0.153	1.28	7.438	51.2
Code	C++	C++	Matlab	C++	Matlab

Table 1: Comparisons of average run time(seconds per image of rough size 300×400.

3.2. Running Time

In Table 1 we compare the average running time of our approach with some other superpixel (region) based methods. Experiments are taken on a machine with an Intel 2.7GHz CPU and 2GB RAM.

We apply the *Toolbox Graph* to compute the shortest path. In our method, graph construction and shortest path computation only cost about 10% (0.14s) of the whole time, with superpixels generation and MDN-based Gaussians spending 40% respectively. The GS method is slower because it uses a large number of superpixels (1500 for average). The CA method is much slower as it requires an exhaustive nearest-neighbor searching. Considering the quality of saliency maps, our approach is also very efficient for many applications.

4. CONCLUSION

This paper proposes a novel and efficient method for salient object detection that makes uses of boundary prior and foci of attention. With the boundary prior, the background is greatly suppressed by a combination strategy of four saliency maps while salient objects are still kept highlighted. In addition, foci of attention based on maximal deviation from norm (MDN) are incorporated to weight the saliency and enhance the quality of the final saliency map. Both evaluation and comparison results show the effectiveness of the proposed method.

Acknowledgements. We thank the anonymous reviewers for their valuable suggestions. This research is partly supported by NSFC, China (No: 61273258, 61375048), Ph.D. Programs Foundation of Ministry of Education of China (No.20120073110018).

5. REFERENCES

[1] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, pp. 1254–1259, 1998.

[2] M. Cheng, G. Zhang, N. Mitra, X. Huang, and S. Hu, "Global contrast based salient region detection," in *proc. CVPR*, 2011, pp. 409–416.

[3] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *proc. CVPR*, 2009, pp. 1597–1604.

[4] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *proc. CVPR*, 2012, pp. 733–740.

[5] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *proc. ECCV*, 2012, pp. 29–42.

[6] K. Fu, C. Gong, J. Yang, and Y. Zhou, "Salient object detection via color contrast and color distribution," in *proc. ACCV*, 2012, pp. 111–122.

[7] Z. Mao, Y. Zhang, K. Gao, and D. Zhang, "A method for detecting salient regions using integrated features," in *proc. ACM MM*, 2012, pp. 745–748.

[8] Z. Wang and B. Li, "A two-stage approach to saliency detection in images," in *proc. ICASSP*, 2008, pp. 965–968.

[9] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 10, pp. 1915–1926, 2012.

[10] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *proc. ICCV*, 2009, pp. 2106–2113.

[11] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li, "Automatic salient object segmentation based on context and shape prior.," in *proc. BMVC*, 2011, pp. 7–18.

[12] H. Yeh and C. Chen, "From rareness to compactness: Contrast-aware image saliency detection," in *proc. ICIP*, 2012, pp. 1077–1080.

[13] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels," *École Polytechnique Fédéral de Lausssanne (EPFL), Tech. Rep*, vol. 149300, 2010.

[14] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.

[15] K. Koffka, "Principles of gestalt psychology," pp. 2–4, 1935.