SALIENCY DETECTION BASED ON INTEGRATION OF CENTRAL BIAS, REWEIGHTING AND MULTI-SCALE FOR SUPERPIXELS

Xiaoling Hu, Wenming Yang, Fei Zhou, Qingmin Liao*

Tsinghua University,China
Shenzhen Key Lab. of Information Sci&Tech/Shenzhen Engineering Lab. of IS&DRM
Department of Electronic Engineering/Graduate School at Shenzhen

ABSTRACT

Saliency detection has been a significant problem in computer vision and helpful to object detection. In this paper, we propose a new computational saliency detection model under the Bayesian framework. First, central bias and the reweighting of the salient regions in the convex hull are applied to guide the prior map. Then, multi-scale for superpixels is proposed to detect objects with various scales. At last, the Bayes formula is adopted to obtain the final saliency map. Experimental results on a standard database show that the proposed model outperforms state-of-the-art methods.

Index Terms— Saliency detection, Bayesian framework, central bias, reweighting, multi-scale

1. INTRODUCTION

Human vision system can quickly pick out interesting parts from a complicated scene. Similar vision system also enables animals to efficiently find prey and avoid predators, which is important for them to survive. Due to this, a lot of researchers have made efforts to explore the mechanism of attention. Salient detection aims to find the parts of an image that attract the most attention, and it can be applied to many research areas in image processing, such as image segmentation [1], object recognition [2], content based image retrieval [3] and so on.

Generally, saliency detection models are categorized as bottom-up and top-down approaches. Bottom-up methods [4, 5, 6] are data-driven, fast and pre-attentive, while top-down methods [7, 8] are goal-driven, slow and entail supervised learning with class labels.

Over the past decade, most researchers have paid attention to bottom-up methods. Bottom-up saliency methods mainly measure saliency by calculating contrast (local or global) or rarity of features over the entire image, and rely on prior knowledge. Different saliency models use different prior knowledge. In [4], Itti *et al.* focus on color and orientation, and adopt center-surround contrast to find salient regions. Bruce *et al.* [9] compute saliency by self-information measurement based on local contrast. On the contrary, Cheng *et al.* [6] use global contrast based on region contrast to detect visual saliency. In this way, we can usually detect the whole salient object. Perazzi *et al.* [10] improve [6]$'s$ work and propose a linear-time computation strategy. Fourier spectrum is also a novel method used to detect visual saliency [11]. Wei *et al.* [12] use geodesic distance to detect salient objects.

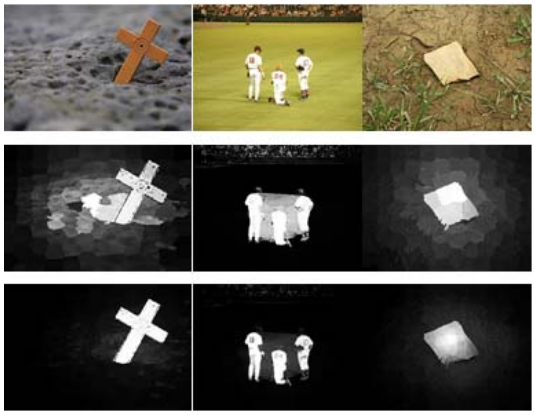

Fig. 1. Saliency maps. From top to bottom: input image, saliency map of [13], our saliency map.

Rahtu *et al.* [14] use a Bayesian model based on sliding window to detect the salient object. Xie *et al.* [13] combine the prior map obtained from convex hull and the observation likelihood in the Bayesian framework to generate the saliency map and achieve the best performance among the state-of-the-art center-surround methods. And the Jing *et al.* [15] promote Xie *et al.*$'s$ model with a boundary based prior map and a soft-segmentation based convex hull. However, there are some shortcomings. First, the salient regions will often be in the center of the image, which is called central bias being neglected. Second, the saliency cluster makes a big differ-

This work was supported by NSFC under Grant No.61471216 and Special Foundation for the Development of Strategic Emerging Industries of Shenzhen under Grant No.YJ20130402145002441 and the China Post-Doctoral Science Foundation under Grant No. 2014T70083.

Email: hxl14@mails.tsinghua.edu.cn, yangelwm@163.com, liaoqm@tsinghua.edu.cn.

ence to the prior map. The regions in the saliency cluster are allocated with the same weight or mainly dependent on the boundary, which will inevitably lead to an inaccurate map. Third, the framework may fail when the scales of the objects vary.

To overcome the above shortcomings, we propose a saliency detection model based on central bias, reweighting the salient regions in the convex hull and multi-scale to improve the performance. The main contribution of our model are twofold: (1) We use central bias and reweighting salient regions in the convex hull to obtain a more accurate prior map. (2) The multi-scale is used to detect salient objects with different scales. Comparison with the other methods shows that our model is effective. Fig.1 shows some comparison samples with [13].

The rest of the paper is organised as follows. In Section 2, we discuss about our model in detail. Experimental results and discussions are presented in Section 3. Section 4 concludes this paper.

2. OUR MODEL

The previous work of saliency detection based on Bayesian law [13, 15] is first presented in Section 2.1, and we will talk about the details of our approach in the following subsections.

2.1. Bayesian Probability

In [13, 15], Bayes formula is adopted to obtain a saliency map. Firstly, Harris Point is used to create a convex hull which locates the object coarsely. Secondly, the regions in the convex hull are divided into two clusters based on superpixels. Then, they take one of the clusters as the preliminary location of the salient region and calculating the observation likelihood. Finally, saliency detection problem is formulated as a Bayesian inference problem for estimating the posterior probability at pixel/location l:

$$p(s|l) = \frac{p(s)p(l|s)}{p(s)p(l|s) + p(b)p(l|b)} \tag{1}$$

$$p(b) = 1 - p(s), \tag{2}$$

where $p(s)$ and $p(b)$ are the prior distribution of the salient regions and background respectively. And $p(l|s)$ and $p(l|b)$ represent the likelihood of observations accordingly.

The method of [13] can detect salient regions effectively. However, as has been shown in the introduction, there are some shortcomings. To improve the method, we propose the central bias and reweighting of the salient regions in the convex hull to gain a better prior saliency map (Section 2.2). Also, multi-scale is used to detect objects with multiple scales (Section 2.3).

2.2. Enhanced Prior Map: Central bias and Reweighting

The saliency cluster makes a big difference to the prior map. In [13], the prior map will be inaccurate when the regions in the saliency cluster have the same weights. In [15], Jing *et al.* propose a boundary based prior map to improve the performance for coarse location of saliency. But in this way, it is the edge that is highlighted rather than the whole region .

It is common sense that the more salient regions should be assigned with higher weights. Therefore a reweighting mechanism is proposed.

In [13, 15], two factors are considered for evaluating the saliency: the color differences between image superpixels, and their spatial distances. With the increasing of the distances, the saliency of the superpixel will be decreased. Furthermore, as stated in [16], the distance of each superpixel from the center of the image should be considered in the evaluation of the saliency. The farther a superpixel is from the center, the less salient it will be accordingly.

Suppose that there are N superpixels in the convex hull. By integrating the elements of color difference, spatial distance and central bias, the preliminary saliency of the superpixel i is defined as follows:

$$s_i{}^0 = \frac{1}{N}(sal_i{}^0 + \frac{1}{N-1}\sum_{j=1}^{N}\delta(j,i)sal_j{}^0), \tag{3}$$

where the function δ is a $Kronecker\ delta$ function, and $sal_i{}^0$ is defined as:

$$sal_i{}^0 = \sum_{j=1,j\neq i}^{N}\omega_1(i)\frac{\omega_2(j)}{d_c(r_j,r_i) + \lambda d_s(r_j,r_i)}, \tag{4}$$

where $d_c(r_j,r_i)$ and $d_s(r_j,r_i)$ represent the color and spatial distances between the j-th and the i-th superpixel, and λ is a tradeoff between color distance and spatial distance. $\omega_1(i)$ is the term we proposed according to the central bias (defined as Eq. 5) and $\omega_2(j)$ is the weighting term.

$$\omega_1(i) = \exp(-DistToCenter(i)/\theta^2), \tag{5}$$

where $DistToCenter(i)$ is the spatial distance between the center of superpixel i and the center of the original image, and θ controls the intensity of spatial weighting.

In Eq. 4, besides the weight term $\omega_1(i)$ representing the bias to the center of image, $\omega_2(j)$ is the second term indicating the weights of the regions in the convex hull. At first, we set:

$$\omega_2(j) = 1, \tag{6}$$

It is apparent that the value $s_i{}^0$ should be high when the superpixel i tends to be salient. Then, we update $\omega_2(j)$ using $s_j{}^0$. We recalculate the saliency of superpixels as follows:

$$sal_i = \sum_{j=1,j\neq i}^{N}\omega_1(i)\frac{s_j{}^0}{d_c(r_j,r_i) + \lambda d_s(r_j,r_i)}, \tag{7}$$

s_i is obtained from Eq. 3 and Eq. 7 as the final saliency value for superpixel i by replacing ${sal_i}^0$ with sal_i. Then the result s_i is normalized to the range [0,1], and this new value is denoted as $p(s)$ and used as the prior distribution in the Bayesian framework.

However, it can extract more information if multi-scale superpixel based method is performed on the existing result. So we propose a superixel based multi-scale method as follows.

2.3. Multi-scale superpixel based method

Though multi-scale for superpixels has been used in [17], there are some differences in our model. A good saliency map should emphasize the pixels with high responses, while those pixels with low responses should be suppressed.

To account for this property we suppress any saliency value below a certain threshold and strengthen any saliency value above another threshold, and s_r is the saliency value at scale r:

$$s_r = \begin{cases} 0, & s_r \leq \alpha \\ 1, & s_r \geq \beta \end{cases}, \tag{8}$$

where α has a small value, such as 0.05, and β has a big value, which equals $1-\alpha$.

And then we use the average saliency value of several scales as the saliency value of one pixel:

$$s = \frac{1}{R}\sum_{r=1}^{R} s_r, \tag{9}$$

where s is the average saliency value, and R is the number of scales.

2.4. Observation Likelihood

Following the work of [13], we compute the observation likelihood in the convex hull. Suppose the salient regions and the background histograms in the CIELAB color space are S_l, S_a, S_b and B_l, B_a, B_b respectively. The color features of a pixel l are l_l, l_a, l_b. Then the observation likelihood is calculated as:

$$p(l|s) = \prod_{i\in\{l,a,b\}} \frac{S_i(l_i)}{N_S} \tag{10}$$

$$p(l|b) = \prod_{i\in\{l,a,b\}} \frac{B_i(l_i)}{N_B} \tag{11}$$

where we suppose that the features in L,A,B space are independent of each other, and N_S and N_B are the number of bins in and out of the convex hull.

And $p(s), p(l|s), p(l|b)$ can be used to obtain $p(s|l)$ with Eq. 1 and Eq. 2.

3. EXPERIMENTS AND COMPARISON

We evaluate our method quantitatively on the 1000-image public dataset [5], which is a subset of the MSRA dataset. Instead of using a rectangle to bound the salient object, accurate human-marked labels are provided as ground truth in this 1000-image dataset.

As to the parameters setting, we choose to scale the image, and the number of the superpixels are 50, 100, 200, 400, and 800. The tradeoff parameter λ of superpixels is set to 1.0 empirically to balance the color feature and the spatial feature. Besides, we set $\theta^2 = 1000$.

3.1. Evaluation Metrics

We evaluate our method by $precision$, $recall$ and $F-measure$. The $precision$ is the percentage of salient pixels correctly detected to all the pixels of extracted regions. And the $recall$ is the percentage of salient pixels correctly detected to the ground truth. Then, we use the average $precision$ and $recall$ to compute the $F-measure$ as:

$$F_\beta = \frac{(1+\beta^2)\times Precision \times Recall}{\beta^2 \times Precision + Recall} \tag{12}$$

where we set $\beta^2 = 0.3$, following [5].

When comparing with other methods, we follow [5, 6, 10, 18] to use an adaptive binarization threshold T_α to binarize the saliency maps obtained by other methods before calculate F_β. The threshold is set as proportional to the mean saliency of the image consistently:

$$T_\alpha = \frac{2}{W\times H}\sum_{x=1}^{W}\sum_{y=1}^{H} S(x,y) \tag{13}$$

3.2. Effectiveness of Our Mechanism

There are two main improvements compared to the work of [13]. The first one is the use of central bias and reweighting salient regions in the convex hull to obtain a more accurate prior map. The second is the utilization of the multi-scale to improve the accuracy.

3.2.1. Validation of central bias, reweighting and multi-scale

To demonstrate the effectiveness of the central bias, reweighting and multi-scale mechanism, we compare our results with other mechanisms by the $precision$, $recall$ and $F-measure$.

In Table 1, we use XL [13] as baseline and add different ingredients to evaluate performance, in which C, R stand for central bias and reweighting as mentioned in Section 2.2, and M stands for multi-scale as mentioned in Section 2.3.

The $F-measure$ of our approach is 0.8331. When the central bias is neglected, the $F-measure$ decreases from

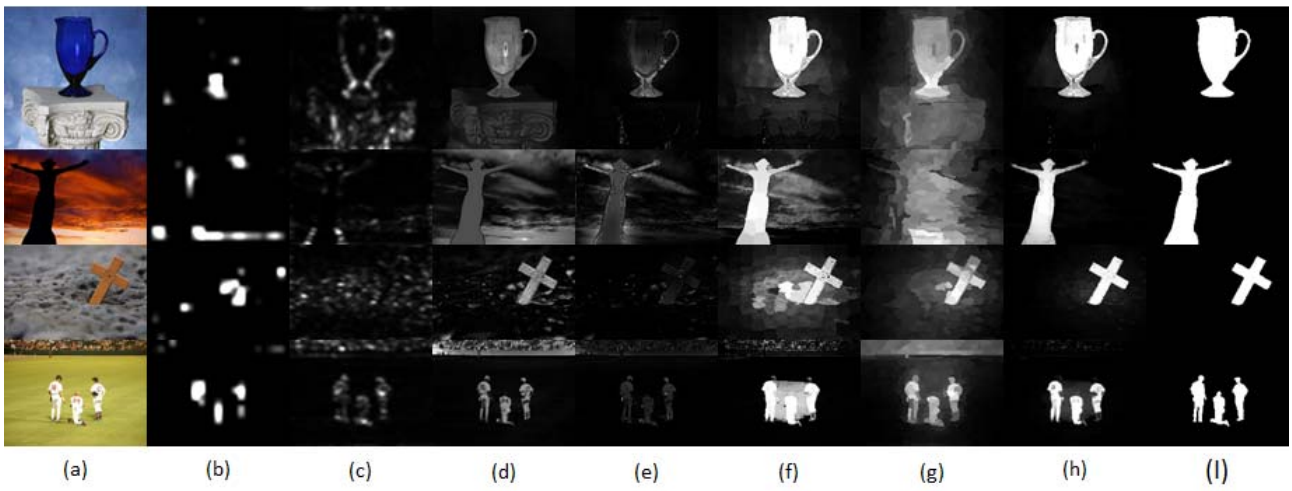

Fig. 2. Saliency detection results of different methods. From left to right:(a)original images, results produced using (b)IT method, (c)SR method, (d)FT method, (e)AC method, (f)XL method, (g)LRM method, (h)our method and (i)ground truth.

Table 1. The $Precision$, $Recall$, $F-measure$ of different mechanisms.

Method	Precision	Recall	F-measure
XL	0.8035	0.8029	0.8034
XL+R+M	0.8362	0.8034	0.8284
XL+C+M	0.8319	0.8053	0.8256
XL+C+R	0.8223	0.8037	0.8180
Ours(XL+C+R+M)	0.8395	0.8125	**0.8331**

0.8331 to 0.8284, the same as reweighting and multi-scale term. The comparison of the different mechanisms in our method validate the effectiveness of the three terms.

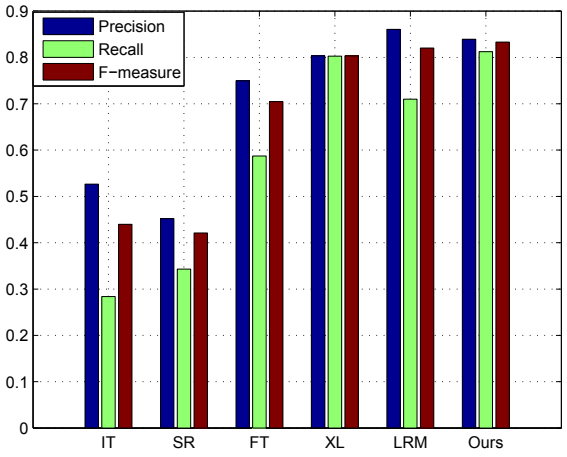

Fig. 3. Average precision, recall and F-measure on the 1000-image dataset. F-measure of the proposed method achieves the best.

3.2.2. Comparison with other methods

On the MSRA-1000 dataset [5], we compare our method with other state-of-the-art approaches, including contrast-based approaches (FT [5]), center-surround (IT [4]), color contrast (AC [19]), Bayesian-based (XL [13]), spectrum-based (SR [11]), and the one with low rank matrix recovery (LRM [18]), most of which were proposed recently. To evaluate these methods, we use the results from the original authors and the codes are available online.

Fig. 2 shows a few visual results of the proposed method. Overall, the results demonstrate that the proposed model outperforms the state-of-the art methods.

We test our method and other methods mentioned above on the 1000 images, and the average $precision$, $recall$ and $F-measure$ are shown in Fig. 3.

4. CONCLUSION

In this paper, we propose an improved central bias, reweighting and multi-scale based Bayesian framework for the saliency detection. The central bias and reweighting of the salient regions in the convex hull are applied to guide the prior map. The multi-scale is used to detect objects with multiple scales. The Bayesian framework is adopted to integrate the prior map and observation likelihood map for a good saliency map. The experiments on the public dataset show that our approach can effectively improve the prior map and detect multi-scale objects.

5. REFERENCES

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

[2] Ueli Rutishauser, Dirk Walther, Christof Koch, and Pietro Perona, "Is bottom-up attention useful for object recognition?," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. IEEE, 2004, vol. 2, pp. II–37.

[3] Tao Chen, Ming-Ming Cheng, Ping Tan, Ariel Shamir, and Shi-Min Hu, "Sketch2photo: internet image montage," *ACM Transactions on Graphics (TOG)*, vol. 28, no. 5, pp. 124, 2009.

[4] Laurent Itti, Christof Koch, and Ernst Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, , no. 11, pp. 1254–1259, 1998.

[5] Ravi Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk, "Frequency-tuned salient region detection," in *Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on*. IEEE, 2009, pp. 1597–1604.

[6] Mingming Cheng, Guoxin Zhang, N Mitra, Xiaolei Huang, and S Hu, "Global contrast based salient region detection, 2011," in *IEEE CVPR*, pp. 409–416.

[7] Tie Liu, Zejian Yuan, Jian Sun, Jingdong Wang, Nanning Zheng, Xiaoou Tang, and Heung-Yeung Shum, "Learning to detect a salient object," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 2, pp. 353–367, 2011.

[8] Jimei Yang and Ming-Hsuan Yang, "Top-down visual saliency via joint crf and dictionary learning," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2296–2303.

[9] Neil Bruce and John Tsotsos, "Saliency based on information maximization," in *Advances in neural information processing systems*, 2005, pp. 155–162.

[10] Federico Perazzi, Philipp Krähenbühl, Yael Pritch, and Alexander Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 733–740.

[11] Xiaodi Hou and Liqing Zhang, "Saliency detection: A spectral residual approach," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.

[12] Yichen Wei, Fang Wen, Wangjiang Zhu, and Jian Sun, "Geodesic saliency using background priors," in *Computer Vision–ECCV 2012*, pp. 29–42. Springer, 2012.

[13] Yulin Xie and Huchuan Lu, "Visual saliency detection based on bayesian model," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011, pp. 645–648.

[14] Esa Rahtu, Juho Kannala, Mikko Salo, and Janne Heikkilä, "Segmenting salient objects from images and videos," in *Computer Vision–ECCV 2010*, pp. 366–379. Springer, 2010.

[15] Jing Sun, Huchuan Lu, and Shifeng Li, "Saliency detection based on integration of boundary and soft-segmentation," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*. IEEE, 2012, pp. 1085–1088.

[16] Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba, "Learning to predict where humans look," in *Computer Vision, 2009 IEEE 12th international conference on*. IEEE, 2009, pp. 2106–2113.

[17] Anurag Singh, Chee-Hung Henry Chu, Michael Pratt, et al., "Multiresolution superpixels for visual saliency detection," in *Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP), 2014 IEEE Symposium on*. IEEE, 2014, pp. 1–8.

[18] Xiaohui Shen and Ying Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 853–860.

[19] Radhakrishna Achanta, Francisco Estrada, Patricia Wils, and Sabine Süsstrunk, "Salient region detection and segmentation," in *Computer Vision Systems*, pp. 66–75. Springer, 2008.