The Open Cybernetics & Systemics Journal

BENTHAM OPEN

Improved Bayesian Saliency Detection Based on BING and Graph Model

Lv Jianyong, Tang Zhenmin and Xu Wei

The Open Cybernetics & Systemics Journal, Volume 9, 2015

ISSN: 1874-110X
DOI: 10.2174/1874110X20150610E008

Article Type: Research Article

Received: April 17, 2015
Revised: April 22, 2015
Accepted: April 27, 2015

Provisional PDF Publication Date: June 10, 2015

Improved Bayesian Saliency Detection Based on BING and Graph Model

Lv Jianyong, Tang Zhenmin and Xu Wei

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu, 210094, P.R. China

Abstract: Saliency detection plays an important role in many computer vision applications. The traditional Bayesian based saliency model using convex hull to circle a coarse salient region, which is inaccurate and unstable. To address this problem, we propose an improved Bayesian framework based saliency method. Firstly, we utilize the BING (Binarized Normed Gradients) method to generate the coarse conspicuity map. Then, we construct a graph model after SLIC superpixel image abstraction, to refine the initial conspicuity map. This is followed by the spatial information based weighting, to produce the final prior map. Secondly, after adaptive threshold, the observation likelihood map is computed by color histogram. Finally, these two maps are combined through Bayesian formula. Experimental results on two benchmark datasets MSRA-1000 and SOD show that our improved method is superior to 13 state-of-the-art alternatives, especially the previous Bayesian saliency models.

Keywords: saliency detection, improved Bayesian framework, BING, graph model, spatial prior.

1. INTRODUCTION

Human visual system (HVS) has the powerful capability to automatically identify the potential most interesting region in complex scene. How to simulate the mechanism of HVS with a computer, has been investigated by experts from multiple research field, including neuroscience, psychology and computer vision. A variety of computational saliency methods are proposed to effectively detect the salient region which attracts humans' attention. As the saliency results can be applied to many tasks, such as object detection, image retargeting, video summarization, etc, saliency detection has become an active topic in recent years [1].

Visual saliency analysis can be divided into two different directions: eye fixation prediction [2] and salient object detection [3]. The former aims to predict a few human visual attention locations and the latter focuses on detecting the whole meaningful generic object. In this paper, we do research on the pure computational, data-driven, bottom-up salient object detection methods [4].

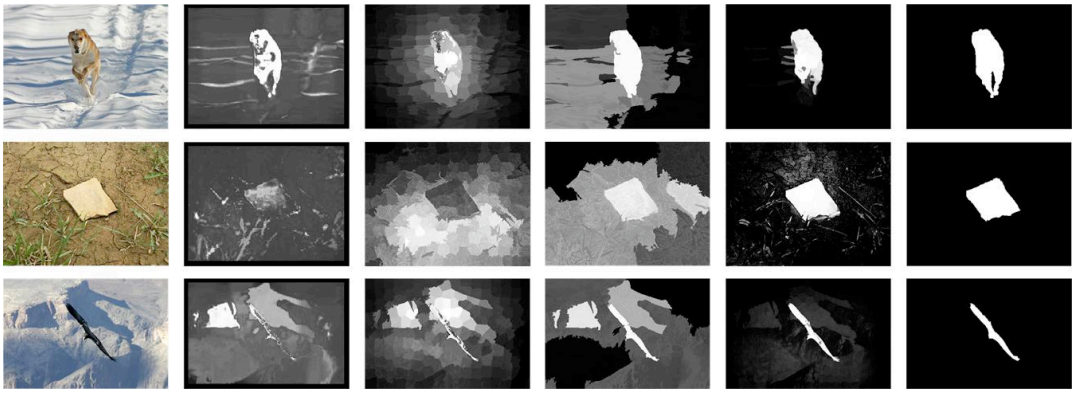
Some local contrast based saliency models estimate the saliency in a particular local region. As the most important work in early stage, Itti [5] et al. utilized center-surrounded differences across multi-scale intensity, color and orientation features to define saliency. AC method [6] directly computed the color difference between the inner and outer local window. After extracting center-surround histogram, multi-scale contrast and spatial distribution, Liu et al. [7] learned a conditional random field to find salient object. Considering the global clue, CA method [8] model saliency by computing the appearance difference between a particular patch and its most similar K ones. However, these methods may produce undesired high salient value near high-contrast edge.

Global methods measure the saliency all over the entire image by exploiting the color statistics, uniqueness, etc. the frequency-tuned method [9] estimates the pixel saliency by its color difference from the average image color. Based on the sparse color histogram, Cheng et al. [4] proposed a regional contrast based salient region detection model (RC). This method is promoted by saliency filter [10], which formulates saliency using N-D Gaussian filters. Then, Cheng et al. [11] used the soft image abstraction and color spatial distribution to improve the saliency results of RC.

Different from the global color contrast method, Margolin et al. [12] utilized the principal component analysis (PCA) to compare the main distinctiveness of image patches. Some graphical saliency models achieve best performance among the traditional global methods [13-14]. Yan et al. [13] constructed a hierarchical graph model and integrated the single-layer saliency cues through a energy minimization function. Yang et al. [14] detected saliency via the graph-based manifold ranking, which considering both foreground and background cues. Furthermore, from the perspective of potential background content, Wei et al. [15] used geodesic distance with respect to background priors to define saliency. More recently, Zhu et al. [16] proposed a robust background measure called boundary connectivity to optimize the saliency results.

The most related to our model is the Bayesian framework based saliency method. Rathu et al [17] presented a sliding window to measure the observation likelihood and set the prior probability as a constant empirically. It is inaccurate and fragile. To tackle with this problem, Xie and Lu [18-19] utilized a coarse-to-fine strategy, which relies on the convex hull to circle a rough foreground region. However, if the initial convex hull is not correct as expected, some background will be assigned the similar salient value to the

*Address correspondence to this author at School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu, 210094, P.R.China; Tel: +86 25 84315660-13;
E-mail: Lv_jy@126.com



(a) Org. (b) RA [17] (c) XL [18] (d) BS [19] (e)Our (f) GT

Fig. (1). Some representative saliency maps of different methods. Org. stands for the original image, GT is the Ground Truth annotated by human. Note that our saliency maps outperform three previous Bayesian saliency models.

real object. Sun et al. [20] improved above method through boundary and soft-segment, but still has high dependence on the convex hull. Different from the rational assumption of [18-20], the main contribution of this paper is that we propose using the statistical object proposal results of BING method to form the coarse conspicuity map, then optimize this map by superpixel based graphical model and spatial prior to get the final prior map. Some comparison of saliency results are shown in Fig. (**1**). Our method can effectively distinguish the background interference and real object.

2.OUR IMPROVED BAYESIAN SALIENCY METHOD

We describe briefly some previous Bayesian saliency models, firstly. Then, we present the details of our prior map and observation likelihood map construction in the Bayesian framework.

2.1 Previous Methods

In [17-20], they define the posterior probability at each pixel x in the Bayesian framework as the saliency measure:

$$p(s \mid x) = \frac{p(s) \cdot p(x \mid s)}{p(s) \cdot p(x \mid s) + p(b)p(x \mid b)} \qquad (1)$$

$$p(b) = 1 - p(s) \qquad (2)$$

where p(s) and p(b) stand for the prior distribution of the salient region and background, respectively; $p(x|s)$ and $p(x|b)$ represent the corresponding observation likelihood. In [17], $p(s)$ is set to a constant and $p(x|s)$ is computed in a double layer sliding window. To get more accurate prior estimation, The methods of [18-20] extract Harris points and circle them to get a convex hull as the coarse saliency region. But these methods still fail in some cases (See Fig. (**1**) (b), (c)).

To overcome the intrinsic shortcoming of convex hull, we present the BING based Bayesian saliency model. The difference of the flowchart between [19] (most representative among Bayesian saliency models) and our method is shown in Fig. (**2**).

2.2 Improved Prior Map Estimation

We use the statistical results of BING method to get the prior map. The BING method is proposed by Cheng et al. [21] to generate the object proposals with image window. Its main theoretical basis is that objects are stand-along things with well-defined closed boundaries, which also satisfies the saliency principle. For a 8×8 image window, it extracts the normed gradients (NG) as a 64D feature firstly. Then, to approximate the NG values, it utilizes the top N_g binary bits of the BYTE values, i.e., the 64D NG feature g_l is described by N_g binarized normed gradients (BING) features:

$$g_l = \sum_{k=1}^{N_g} 2^{8-k} b_{k,l} \qquad (3)$$

where l is the location of the window, $b_{k,l} \in \{0,1\}^{8\times 8}$. The fast BING calculation algorithm is referred in [21]. Then, each window should be scored with a linear model $w \in R^{64}$, which can be approximated with a set of basis vectors [22]:

$$w \approx \sum_{j=1}^{N_w} \beta_j \alpha_j \qquad (4)$$

where N_w is the number of basis vectors, $\alpha_j \in \{-1,1\}^{64}$ represents a basis vector, and β_j is the coefficient. α_j can be modified as

$$\alpha_j = \alpha_j^+ - \overline{\alpha_j^+} \qquad (5)$$

where $\alpha_j^+ \in \{0,1\}^{64}$. Then, the score s_l is computed as:

$$s_l = < w, g_l > \approx \sum_{j=1}^{N_w} \beta_j \sum_{k=1}^{N_g} C_{j,k} \qquad (6)$$

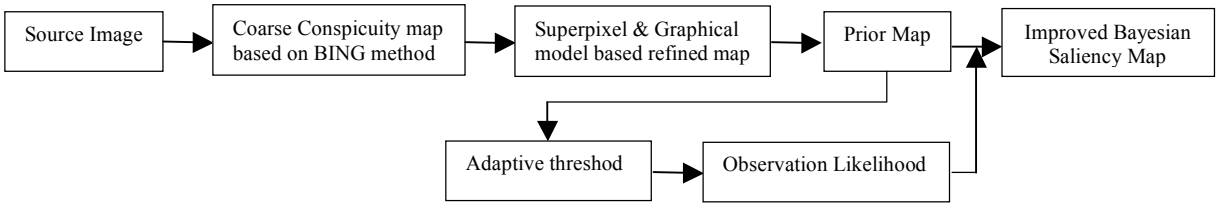
where $C_{j,k} = 2^{8-k}(2 < \alpha_j^+, b_{k,l} > - | b_{k,l} |)$.

The BING method is very efficient to provide object proposals with a small set of category-independent image windows. But due to the rectangle shape constraint of image window with different sizes, this method can't distinguish the whole meaningful object from complex background precisely. Here, we use it to construct the initial coarse conspicuity map.

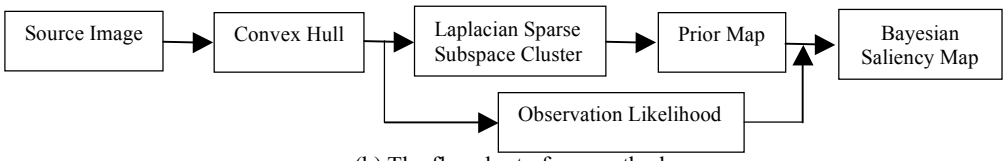
Assume there are N proposal image windows sampled in image I using BING, and the corresponding objectness score(or probability) of each window w_i is $p(w_i)$. For a pixel x, if it locates in w_i, we assign its probability $p(x,w_i) = p(w_i)$ to indicate its objectness, otherwise, $p(x,w_i)=0$. Then, we overlap all the N windows to obtain the probability of x:

$$P_o(x) = \frac{1}{N} \sum_{i=1}^{N} p(x, w_i) \qquad (7)$$

We set N=1000 according to the outstanding performance stated in [21]. The probabilities of all pixels constitute the coarse conspicuity map. Fig. (**3**)(e) shows the visual effect.



(a) The flowchart of the previous Bayesian method in [20]



(b) The flowchart of our method

Fig. (2). The comparison of different Bayesian saliency models

Its location of potential salient region is more accurate than the convex hull (See Fig. (**3**)(c)).

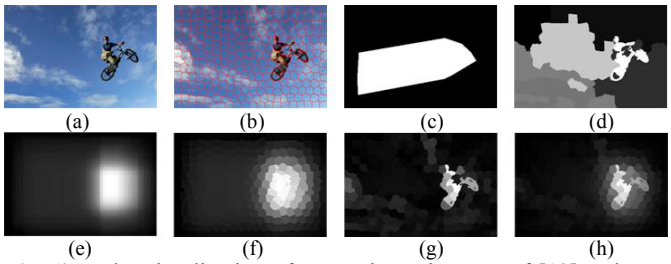


Fig. (3). The visualization of generating prior map of [19] and our method. (a) is the input image, (b) is the superpixel decomposition results, (c) is the convex hull, (d) is the prior map of [19], (e) is the coarse conspicuity map based on BING, (f) is the refined conspicuity map via superpixel, (g) is the graph based color contrast map and (h) is our prior map.

To generate more perceptually accurate conspicuity region, we use SLIC method [23] to decompose an image into superpixels, which can treat large-scale homogeneous pixels with similar features as a unit and well preserve the global object boundaries. If the image is over-segmented into M superpixels and the i-th superpixel is sp_i, we formula the refined conspicuity value of sp_i as:

$$C_i = \frac{1}{N_i}\sum_{k=1}^{N_i} P_o(x_k),\ x_k \in sp_i \qquad (8)$$

where N_i is the number of pixels included in sp_i. Compute all the M refined conspicuity values and normalize them to [0,1]. (See Fig. (**3**)(f)).

The global saliency clue should be involved to improve the distinction between real salient regions and background. But different from the strategy of [19], which uses Laplacian sparse subspace cluster to aggregate all the superpixels into several color clusters and compute the intersection area with respect to the convex hull as the final prior map, we employ the graphical model and boundary color contrast to enhance our conspicuity map.

We construct an image graph $G = (V, E)$, where nodes V stand for the superpixles and edges E represent the links between different nodes. Motivated by [14], we define this graph sparsely connected—each node not only connects to its neighboring nodes, but also the nodes that share common boundaries with its neighborhood (See Fig. (**4**)). Then, for a node i and its adjacent node j in the neighborhood region R, the weigh w_{ij} of the edge e_{ij} is defined as

$$w_{ij} = e^{-\|c_i - c_j\|/\sigma^2},\ j \in R \qquad (9)$$

where c_i and c_j represent the average color of sp_i and sp_j in CIELab space, respectively.σ^2=0.1 in our experiments.

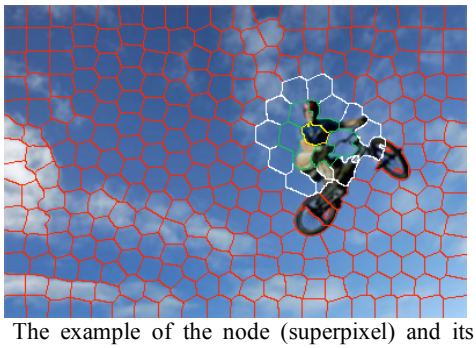


Fig. (4). The example of the node (superpixel) and its sparsely connected local region in graph model. The center node (yellow color) connects to its directly adjacent node (green color), and the nodes that share common boundaries with its neighborhood (white color)

Inspired by [14,15], we also assume that the majority of nodes along image boundaries belong to real background and utilize the "one-dimension rule" to deal with the situation that part of salient object may touch the image boundaries. But different from using manifold ranking [14] or geodesic distance [15], we just adopt color contrast to measure the difference.

If there are N background nodes, b represents the b-th one, the background contrast of node i is defined as the summation of its K minimum color distance with respect to the background nodes in CIELab space:

$$BC_i = \min\{\sum_{b=1}^{K} \| C_i - C_b \|\},\ b \in N \qquad (10)$$

BC_i represents the distinctiveness between node i and the estimated background. Intuitively, the real object is usually consisted of homogeneous superpixels. To get more stable and accurate results, we consider the neighborhood influence of node i in the sparsely connected region R to get the final color contrast (CC):

$$CC_i = \lambda \cdot BC_i + (1-\lambda) \cdot \max\{w_{ij} BC_j\} \quad j \in R \qquad (11)$$

Where λ is the balance parameter. In our experiment , λ=0.4. The graph based color contrast map is shown in Fig. (**3**) (g).

Integrate (8) and (11), the final refined conspicuity value of sp_i is:

$$C'_i = C_i + CC_i \qquad (12)$$

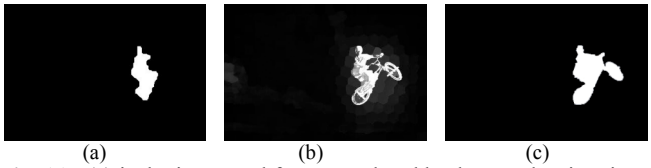
The spatial information has been widely used in many state-of-the-art saliency models [3,7,11,12]. According to the phenomenon call "center-bias", which originates from the humans' habit of placing the object in the middle of image when photograph, we optimize the per-pixel conspicuity map to get the prior map. The value of pixel x in the prior map $p(s)$ is computed by:

$$p(s) = C'(x) \cdot e^{-\alpha \|l_x - l_c\|} \qquad (13)$$

where l_x is the location of x, and l_c is the center location of the prior map. $\alpha = 1/\max(H,W)$ controls the spatial effect. H and W are the height and width of image, respectively. See Fig. (**3**) (h), our prior map highlights the real object while the previous Bayesian method in [19] makes part of redundant background salient (Shown in Fig. (**3**) (d)).

2.3 Enhanced Observation Likelihood

Obviously, a more accurate prior map can greatly benefit the computation of observation likelihood. Different from [19], which directly supposes all the pixels inside the convex hull to be foreground and these outer pixels to be background (see Fig. (**3**) (c)), we use the adaptive threshold strategy described in [24] to binarize the prior map (see Fig. (**3**) (h)). As shown in Fig. (**5**) (a), our improved estimation result is more consistant with the human visual observation result (Fig. (**5**) (c)) compared with the convex hull (Fig. (**3**)(c)).



(a) (b) (c)

Fig. (5). (a) is the improved foreground and background estimation result after binarization, (b) is the saliency map of our method, (c) is the Ground Truth labeled by human

Then we also utilize the center-surround way [19] to acquire the observation likelihood based on the improved foreground and background estimation result. In CIELab color space, we define the color histogram of F (Foreground) and B (Background) as $\{F_L, F_a, F_b\}$ and $\{B_L, B_a, b_b\}$ respectively. And the corresponding numbers of bins in F and B are N_F and N_B. The values of pixel x in these two histograms are $F_i(x)$ and $B_i(x)$, $i \in \{L, a, b\}$. Assume the histograms of L, a, b are independent, the observation likelihood of pixel x can be computed by:

$$p(x \mid s) = \prod_{i \in \{L,a,b\}} \frac{F_i(x)}{N_F} \qquad (14)$$

$$p(x \mid b) = \prod_{i \in \{L,a,b\}} \frac{B_i(x)}{N_B} \qquad (15)$$

Take $p(s)$ of (13), $p(x|s)$ of (14) and $p(x|b)$ of (15) into (1), we can produce the improved Bayesian saliency map(Fig. (**5**)(b)).

3. EXPERIMENTS

We evaluate the proposed method on two benchmark datasets: the widely used MSRA-1000 dataset [9] and the most difficult SOD dataset [15] with human annotated Ground Truth (GT). 13 state-of-the-art methods are compared to complete the qualitative and quantitative analysis, including: RC [4], Itti [5], AC [6], FT [9], SF [10], GC [11], PCAS [12], GS-SP [15], RA [17], XL [18], BS [19], CB [24], LR [25]. As to the parameters setting, we decompose each image into 200 superpixels for efficiency. In (10), we set K=10 for in the following experiments.

3.1 Evaluation on MSRA-1000 Dataset

The MSRA-1000 dataset is a subset of the MSRA dataset [7] and it is a commonly used dataset for evaluating the salient object detection methods. The qualitative comparison of some representative saliency maps generated from different methods are shown in Fig. (**6**) (Org. means the original image). It is obvious that our proposed method can highlight the real salient objects (the white pixels in GT) more accurately and uniformly than the alternatives. Note that our improved Bayesian saliency model apparently outperforms RA, XL and BS (three previous Bayesian saliency method). RA tends to produce higher salient values in some background regions with high local contrast. In some cases, the salient values of real object and part of background are indistinguishable (the second image of RA in Fig. (**6**)). XL and BS are highly dependent on the convex hull, which may divide a lot of redundant background into salient region improperly, leading to an inaccurate map inevitably (the first and second images of XL and BS in Fig. (**6**)). Even in the extreme case, convex hull wrongly circles part of background as the real object, which may greatly enhance the undesired background while ignoring the whole object in saliency map (the third image of XL and BS in Fig. (**6**)). These shortcomings will not appear in our method.

We adopt the precision-recall (P-R) curves generated by varying the fixed threshold from 0 to 255 and computing the corresponding precision and recall with respect to the GT, as the first quantitative evaluation [9]. Precision stands for the percentage of salient pixels correctly assigned with respect to all detected salient pixels, while recall means the percentage of correct salient pixels with respect to the GT. As shown in Fig. (**7**), our method outperforms others when the recall value is under 0.85. when the recall is in the range [0.85, 1], the precision of our method is slightly lower than GS-SP, but still higher than most compared methods. The highest precision value of our method is 0.95, superior to GS-SP (0.91). Compared to the three Bayesian saliency model: RA, XL and BS, the precision of our improved model is much higher.

As the overall performance, the F-measure at a fixed threshold is computed by

$$F - measure = \frac{(1+\beta^2)\,\mathrm{Pr}\,ecision \times \mathrm{Re}\,call}{\beta^2\,\mathrm{Pr}\,ecision + \mathrm{Re}\,call} \qquad (16)$$

where ß 2=0.3 to assign precision higher weight according to [9-11]. We calculate all the F-measures on the thresholds

ranging from 0 to 255 just like getting the P-R curves [20]. As shown in Fig. (**8**), our method has the highest F-measure 0.871 when T=183 (CB method [24] has the second highest

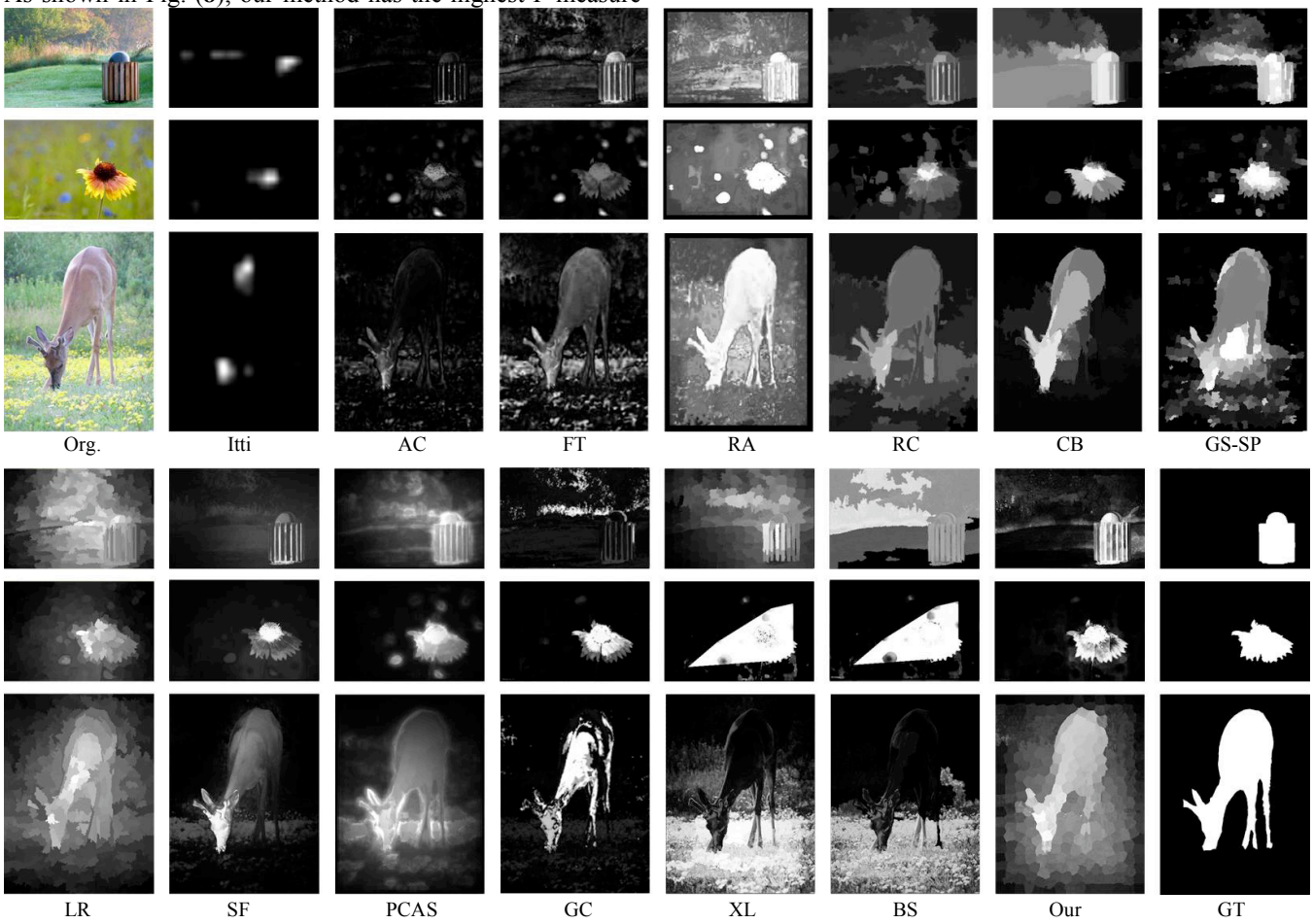


Fig. (6). The visual comparison of our saliency maps with 13 state-of-the-art methods.

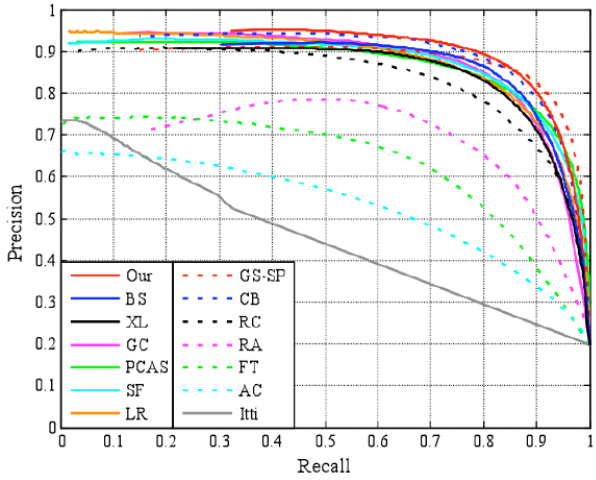


Fig. (7). The P-R Curves of different methods

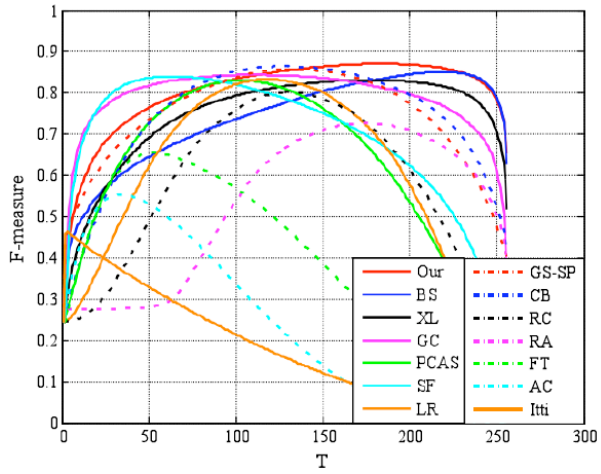


Fig. (8). The F-measure vs T (threshold) Curves of different methods

F-measure 0.866 when T=125) and it keeps high level when T varies in a wide range [100, 225]. The F-measure of RA (0.727), XL (0.832) and BS (0.852) are much lower than ours, indicating that the proposed method indeed improves the Bayesian Framework.

We also adopt the adaptive threshold strategy explained in [9] which is defined as twice of the mean salient value to get the average precision, recall and F-measure (see Fig. (**9**)). We conclude from the comparisons that: (1) The F-measure of our proposed method (0.836) outperforms 12 state-of-the-art methods except CB (0.861) which employs the global context and shape salient clues. (2) Compared to RA, XL and BS, the precision and F-measure of our method is best, demonstrating the efficiency of our improvement.

The above P-R curve, F-measure and adaptive threshold measure don't consider the number of pixels that is marked as

non-salient region correctly (the black pixels in GT, as shown in Fig. (**6**)). To get more balanced comparison results, we use

$$MAE = \frac{1}{W \cdot H} \sum_{i=1}^{W} \sum_{j=1}^{H} | S(i,j) - GT(i,j) | \qquad (17)$$

where W and H are the width and height of the saliency map. The MAE histograms are shown in Fig. (**10**). Our improved method produces the lowest MAE value 0.094, which proves that it provides a better overall similarity measure between the saliency map and GT, including the foreground (white in GT) and the background (black in GT).

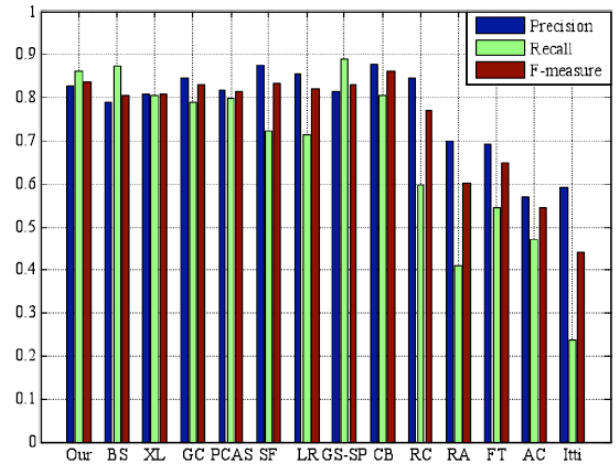


Fig. (9). The Precision, Recall and F-measure after adaptive threshold

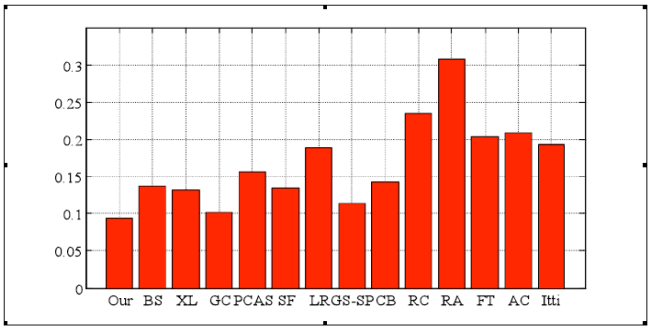


Fig. (10). The MAE histograms of different methods

3.2 Evaluation on SOD Dataset

The SOD dataset is considered as the most difficult and challenging dataset, as it contains 300 images with multiple objects and there are great variations in their scales [15]. Similar to [26], due to the complexity of SOD, we choose the 10 state-of-the-art methods with better performance to compare their P-R curves, F-measure vs T curves, adaptive threshold results and the MAE histograms(as shown in Fig. (**11**)) with our method. These four quantitative performance evaluation results of all the methods are generally not very good.

As observed from Fig. (**11**)(a), The GS-SP method, which presents the SOD dataset and the human labeled Ground Truth, has the best precision and recall rate. Our method produces the second best result on this quantitative indicator. Compared to the previous Bayesian saliency method: RA, XL and BS, our method significantly promotes the precision and recall. When the recall is below 0.1, the precision of our method reach the best value (about 0.8).

Similar to the comparison results of P-R curves, as shown in Fig. (**11**)(b), the highest F-measure of our method (0.616) is only lower than GS-SP (0.650), but still higher than others, especially RA,XL and BS.

After adaptive threshold, the average F-measure of our method (0.569) is lower than PCAS (0.592) and GS-SP (0.632), but higher than RA (0.39), XL (0.519), BS (0.535) (See Fig. (**11**) (c)). Actually, All these F-measures are not good enough, implying that the simple binarization strategy based on the saliency maps cannot accurately extract the real salient object when dealing with complex images.

As shown in Fig. (**11**)(d), the MAE value of our method is 0.276, very close to GC (0.273), PCAS (0.275) and GS-SP (0.280). It is apparently superior to RA (0.360), XL (0.320) and BS (0.308).

All of the above four quantitative comparisons demonstrate that our method has better overall performance than most state-of-the-art methods, including three previous Bayesian saliency models, but still need further improvement when confronting the difficult images with multiple objects.

Several representative saliency maps of different methods are shown in Fig. (**12**). Take the large-scale object (the first image in Fig. (**12**)) for an example, our method produces the best result, which is very similar to GT. Observe the three previous Bayesian saliency model: RA only highlights a small part of salient object; XL wrongly assigns part of background with high salient value; BS generates lower salient value in some important part of real object. For the low contrast image and the image with two different objects (the second and third image in Fig. (**12**)), our method can also produce excellent visualization results, verifying the efficiency of our improvement.

3.3 Failure Cases

Our method is based on the Bayesian model. Although we make great improvement to get more accurate prior map and observation likelihood in the Bayesian framework, but there still exist some images which don't accord with our principle. Some failure cases are shown in Fig. (**13**). When the color of real salient object is very close to most background, especially existing multiple objects with similar color to background, our method will fail. In these cases, RA, XL and BS generate worse results.

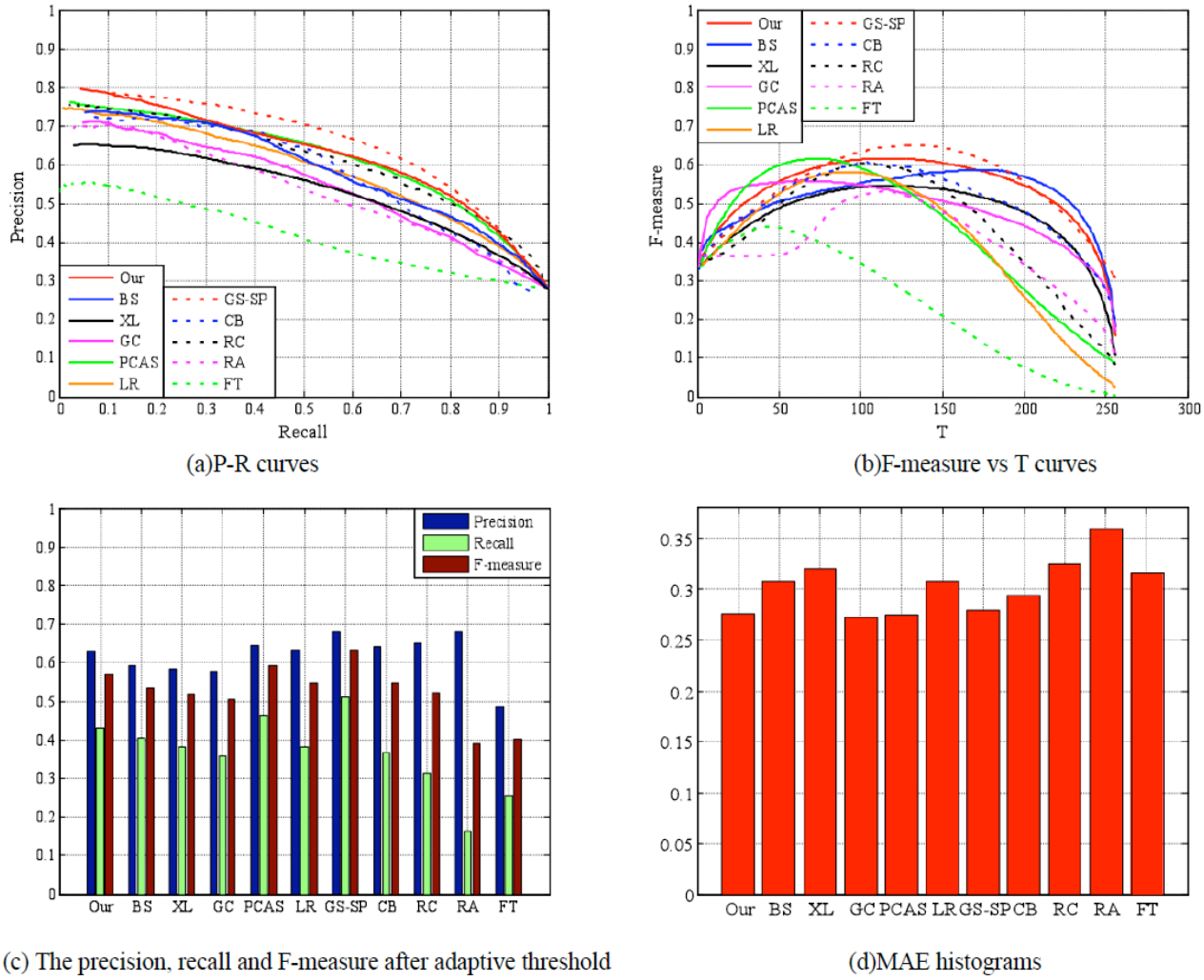
3.4 Running Time Comparison

Finally, we compare the average running time of these methods on MSRA-1000 and SOD datasets. We do experiments using a computer with Inter(R) Core(TM)i5-2410M 2.8GHz CPU and 8GB RAM. Our method is implemented by Matlab and the average running time is 3.32 s for an image. It's not very fast. However, considering the evaluation of quantitative performance, our method still has advantage over these state-of-the-art methods.

Table 1. Comparison of average running times

Method	Our	BS	XL	GC	PCAS	SF	LR
Time(s)	3.32	156.45	2.18	0.09	6.17	0.15	22.34

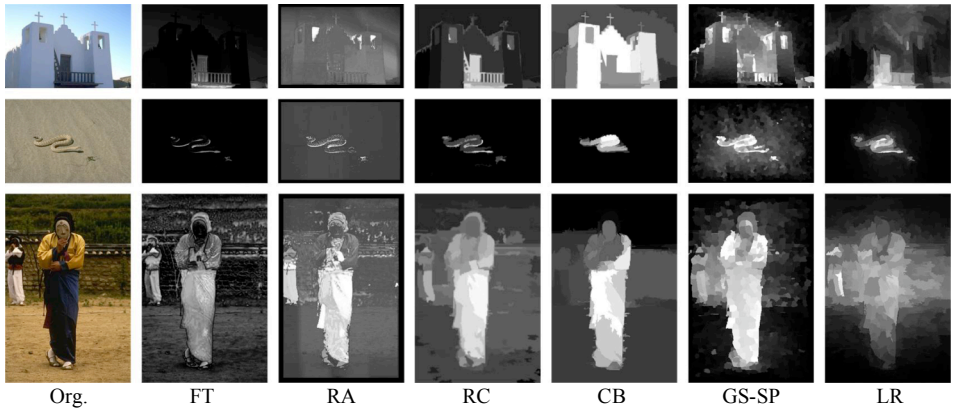
Method	GS-SP	CB	RC	RA	FT	AC	Itti
Time(s)	7.39	2.75	0.13	9.48	0.18	0.06	0.33



(a)P-R curves (b)F-measure vs T curves

(c) The precision, recall and F-measure after adaptive threshold (d)MAE histograms

Fig. (11). The four quantitative performance evaluations on SOD dataset of our method and 10 state-of-the-art methods



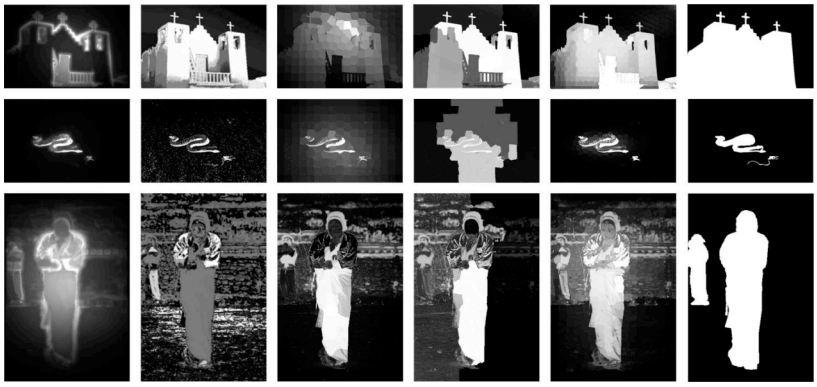


Fig. (12). The visual comparison of 10 previous approach and our method

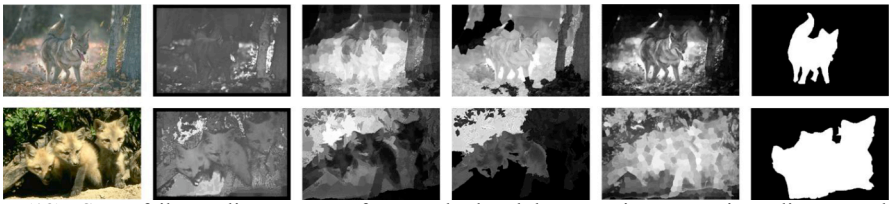


Fig. (13). Some failure saliency maps of our method and three previous Bayesian saliency models

4. CONCLUSIONS

In this paper, we propose an improved Bayesian saliency model through the objectness proposal method—BING and graph based boundary color contrast measure. The statistical probability results of BING can provide the coarse location of potential object more accurately and robustly compared with the convex hull based Bayesian saliency model. Then, the graph based color contrast measure has greatly enhanced the prior map, which also benefit the observation likelihood. After the integration via Bayesian formula, we get the final saliency map with better performance. The experimental results indicate that our method outperforms all 13 sate-of-the-art methods in qualitative and quantitative evaluations, especially these methods based on the similar Bayesian framework ― RA,XL and BS. In future work, we will investigate discriminative salient features to estimate the location of potential object more accurately.

CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This research was supported by the Natural Science Foundation of China (NSFC), No. 61473154.

REFERENCES

[1]A Borji, L Itti, "State-of-the-art in visual attention modeling", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 35, No. 1,pp. 185-207, 2013.

[2]A Borji, D N Sihite, L Itti, "Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study". IEEE Transactions on Image Processing, Vol. 22, No. 1, pp. 55-69, 2013.

[3]A Borji, D N Sihite, L Itti, "Salient object detection: a benchmark", Proc. European Conference on Computer Vision, Springer Press, 2012, pp. 414-429.

[4]M Cheng, G Zhang, N J Mitra, X Huang, S Hu, "Global contrast baed salient region detection", Proc. IEEE International Conference on Computer Vision and Pattern Recognition. IEEE Press, 2011, pp. 409-416.

[5]L Itti, C Koch, E Niebur, "A model of saliency-based visual attention for rapid scene analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 11,pp. 1254-1259, 1998.

[6]R Achanta, F Estrada, P Wils, S Süsstrunk, "Salient region detection and segmentation", Proc. IEEE International Conference on Computer Vision Systems, Springer Press, 2008, pp.66-75.

[7]T Liu, Z Yuan, J Sun, J Wang, N Zheng, X Tang, et al, "Learning to detect a salient object", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, No. 2,pp. 353-367, 2011.

[8]S Goferman, L Zelnik-Manor, A Tal, "Content-aware saliency detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 34, No. 10,pp. 1915-1926, 2012.

[9]R Achanta, S Hemami, F Estrada, S Süsstrunk, "Frequency-tuned salient region detection", Proc. IEEE International Conference on Computer Vision and Pattern Recognition, IEEE Press, 2009, pp.1597-1604.

[10]F Perazzi, P Krahenbuhl, Y Pritch, A Hornung, "Saliency filters: Contrast based filtering for salient region detection", Proc. IEEE Conference on Computer Vision and Pattern Recognition, IEEE Press, 2012, pp. 733-740.

[11]M Cheng, W Jonathan, W Lin, Z Shuai, V Vibhav, C Nigel, "Efficient salient region detection with soft iamge abstraction", Proc. IEEE International Conference on Computer Vision, IEEE Press, 2013, pp. 1529-1536.

[12]R Margolin, A Tal, L Zelink-Manor, "What makes a patch distinct?", Proc. IEEE Conference on Computer Vision and Pattern Recognition, IEEE Press, 2013, pp. 1139-1146.

[13]Q Yan, L Xu, J P Shi, J Y Jia, "Hierarchical saliency detection", Proc. IEEE International Conference on Computer Vision and Pattern Recognition, IEEE Press, 2013, pp.1155-1162.

[14]C Yang, L H Zhang, H C Lu, X Ruan, M H Yang, "Saliency detection via graph-based manifold ranking", Proc. IEEE International Conference on Computer Vision and Pattern Recognition, IEEE Press, 2013, pp. 3166-3173.

[15]Y Wei, F Wen, W Zhu, S Jian, "Geodesic saliency using background priors", Proc. European Conference on Computer Vision, Springer Press, 2012, pp. 29-42.

[16]W J Zhu, S Liang, Y C Wei, J Sun, "Saliency optimization from robust background detection", Proc. IEEE International Conference on Computer Vision and Pattern Recognition, IEEE Press, 2014, pp. 2814-2821.

[17]E Rahtu, J Kannala, M Salo, J Heikkila, "Segmenting salient objects from images and videos", Proc. European Conference on Computer Vision, Springer Press, 2010, pp. 366-379.

[18]Y Xie, H C Lu, "Visual saliency detection based on Bayesian model", Proc. IEEE International Conference on Image Processing, IEEE Press, 2011, pp. 645-648.

[19]Y Xie, H Lu, M Yang, "Bayesian saliency via low and mid level cues", IEEE Transactions on Image Processing. Vol. 22, No. 5, pp. 1689-1698, 2013.

[20]J Sun, H C Lu, S F Li, “Saliency detection based on integration of boundary and soft-segmentation”, Proc. IEEE International Conference on Image Processing, IEEE Press, 2012, pp. 1085-1088.
[21]M M Cheng, Z M Zhang, W Y Lin, P Torr, “BING: Binarized normed gradients for objectness estimation at 300fps”, Proc. IEEE Conference on Computer Vision and Pattern Recognition, IEEE Press, 2014, pp. 3286 -3293.
[22]S Hare, A Saffari, P H Torr, “Efficient online structured output learning for keypoint-based object tracking”. Proc. IEEE Conference on Computer Vision and Pattern Recognition, IEEE Press, 2012, pp. 1894 -1901.
[23]R Achanta, A Shaji, K Smith, A Lucchi, P Fua, S Süsstrunk, “SLIC superpixels compared to state-of-the-art”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 34, No. 11, pp. 2274-2282, 2012.
[24]H Jiang, J Wang, Z Yuan, T Liu, N Zheng, S Li, “Automatic salient object segmentation based on context and shape prior”, Proc. British Machine Vision Conference, BMVA Press, 2011, pp.110.1-110.12.
[25]X Shen, Y Wu, “A unified approach to salient object detection via low rank matrix recovery”, Proc. IEEE Conference on Computer Vision and Pattern Recognition, IEEE Press, 2012, pp. 853-860.
[26]B Jiang, L Zhang, H Lu, C Yang, M Yang, “Saliency detection via absorbing markov chain”, Proc. IEEE International Conference on Computer Vision, IEEE Press, 2013, pp. 1665-1672.