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 is used based on convex hull to circle a coarse salient region, which is inaccurate and unstable. To address this problem, this paper proposed an improved Bayesian framework based saliency method. Firstly, the BING (Binarized Normed Gradients) method was utilized to generate the coarse conspicuity map. Then, a graph model was constructed 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 was computed by color histogram. Finally, these two maps were combined using Bayesian formula. Experimental results on two benchmark datasets MSRA-1000 and SOD show that the improved method was superior to 13 state-of-the-art alternatives, especially the previous Bayesian saliency models.

1. INTRODUCTION

Human visual system (HVS) has the dominant capability to automatically identify the potential and the most interesting region in complex scene. How to simulate the mechanism of HVS with a computer, has been investigated by experts of multiple research field, including neuroscience, psychology and computer vision. A variety of computational saliency methods have been 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 a hot 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 locations and the latter focuses on detecting the whole meaningful generic object. In this paper, the pure computational, data-driven, and bottom-up salient object detection methods were researched [4].

Some local contrast based saliency models estimate saliency in a particular local region. As the most important work in the 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 the outer local window. After extracting center-surround histogram, multi-scale contrast and spatial distribution, Liu *et al.* [7] employed a conditional random field to find the salient object. Considering the global clue, the CA method [8] models

saliency by computing difference in the appearance of particular patches and the most similar K ones. However, these methods may produce undesired high salient value near the 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 the 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 the 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 energy minimization function. Yang *et al.* [14] detected saliency *via* the graph-based manifold ranking, which considers both the foreground and background cues. Furthermore, from the perspective of potential background content, Wei *et al.* [15] used geodesic distance with respect to background prior 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 on the saliency method. Rathu *et al.* [17] presented a sliding window to measure the observation likelihood and empirically set the prior probability as a constant, which is inaccurate. To tackle this problem, Xie and Lu [18, 19] uti-

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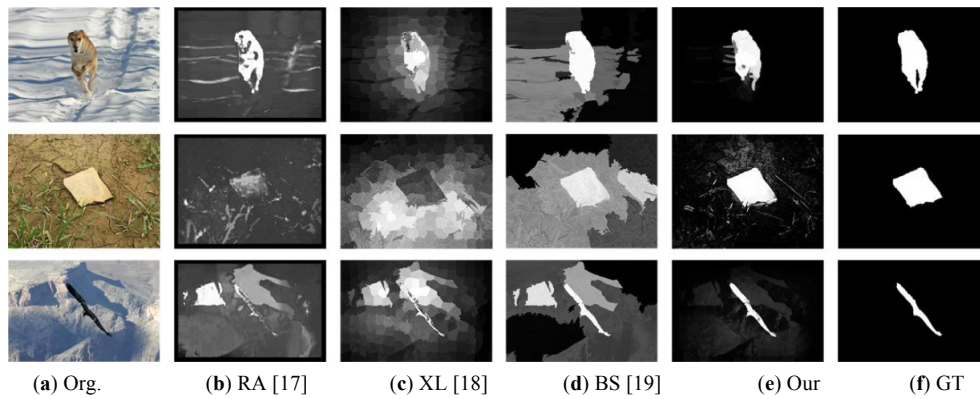
Fig. (1). Some representative saliency maps of different methods. Org. stands for the original image, GT is the Ground Truth annotated by human. Note that our saliency maps outperform three previous Bayesian saliency models.

lized 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 to obtain the similar salient value of the real object. Sun *et al.* [20] improved the 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 it proposed using the results of statistical object of BING method to form the coarse conspicuity map, and then optimized this map by superpixel based on graphical model to obtain the spatial prior to determine the final prior map. The comparison of saliency results is shown in Fig. (**1**). The method proposed in this study can effectively distinguish the background interference and real object.

2. OUR IMPROVED BAYESIAN SALIENCY METHOD

This paper briefly described some previous Bayesian saliency models. Furthermore, the details of the prior map are also presented along with the construction of observation likelihood map I based on the Bayesian framework.

2.1. Previous Methods

In [17-20], the posterior probability is defined 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 prior distribution of the salient region and background, respectively; $p(x|s)$ and $p(x|b)$ represent the corresponding observation likelihood. In [17], $p(s)$ was set as a constant and $p(x|s)$ was computed in a double layer sliding window. For more accurate prior estimation, using the methods of [18-20], Harris points are extracted and circled to form a convex hull as the coarse saliency region. But these methods still fail in some cases (See Figs. (**1b** and **c**).

To overcome the intrinsic shortcoming of convex hull, this paper presented 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

This paper used the statistical results of BING method to construct the prior map. The BING method was proposed by Cheng *et al.* [21] to generate the object proposals with the image window. Its main theoretical basis is that the objects are stand-alone things with well-defined closed boundaries, which also satisfies the saliency principle. For an 8×8 image window, it firstly extracts the normed gradients (NG) as a 64D feature. Then, to obtain the approximate 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:

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

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]:

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

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

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

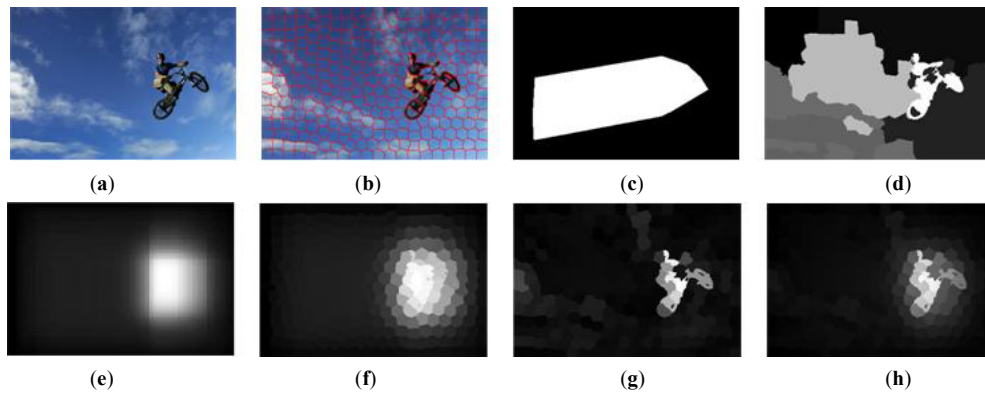Where $\alpha_j^+ \in \{0,1\}^{64}$. Then, the score s_l is computed as:

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

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 constraint of the rectangular shape of the image window with different sizes, this method cannot precisely distinguish the whole meaningful object from the complex background. Here, this was employed to construct the initial coarse conspicuity map.

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

(**b**) The flowchart of our method
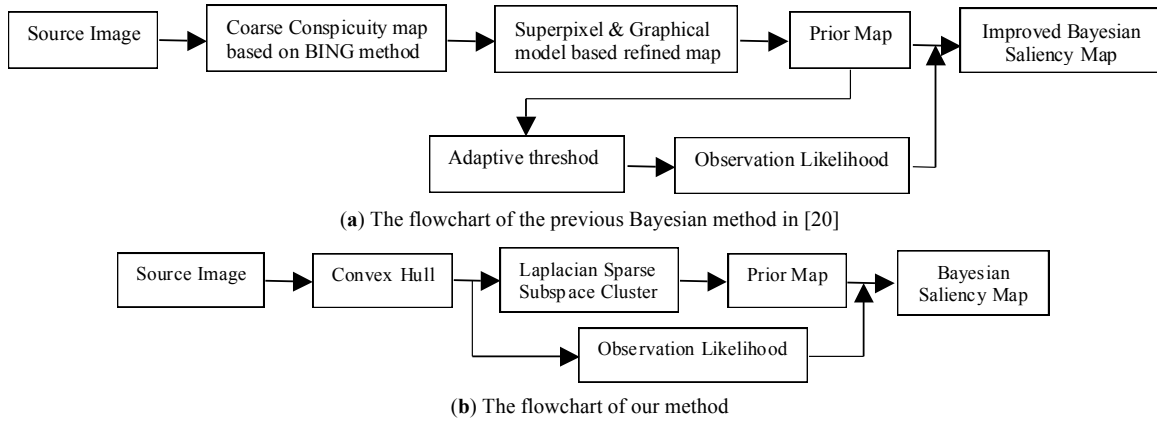
Fig. (2). The comparison of different Bayesian saliency models.

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.

It was assumed that there were *N* proposed image windows sampled in the image *I* using BING, and the corresponding objectness score (or probability) of each window w_i was $p(w_i)$. For a pixel *x*, when it was located in the w_i, its probability was assigned as $p(x,w_i)= p(w_i)$ to indicate its objectness, otherwise, $p(x,w_i)=0$. After this, all the *N* windows were overlapped to obtain the probability of *x*:

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

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

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

To generate more perceptually accurate conspicuity region, the paper used 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, the formula is refined with conspicuity value of sp_i as:

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

Where N_i is the number of pixels included in sp_i. Following this, all the *M* refined conspicuity values were computed and normalized to [0,1]. (See Fig. (**3f**)).

The global saliency clue should be used to improve the distinction between the real salient regions and background. But different from the strategy mentioned [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, this study employed the graphical model and boundary color contrast to enhance the conspicuity map.

The paper constructed an image graph $G = (V, E)$, where nodes *V* stand for the superpixels and edges *E* represent the links between different nodes. Motivated by a study[14], this graph was defined as sparsely connected, where—each node not only connects to its neighboring nodes, but also the nodes that share common boundaries with their neighboring nodes (See Fig. (**4**)). 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:

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

Where, c_i and c_j represent the average color of sp_i and sp_j in the CIELab space, respectively. $\sigma^2 = 0.1$ in the experiments.
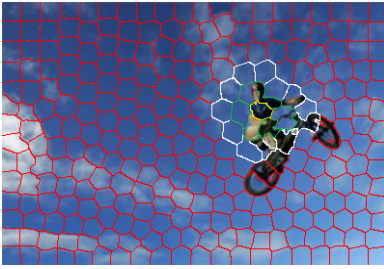
Fig. (4). The example of the node (superpixel) and its sparsely connected local region in the graph model. The center node (yellow color) connects to its directly adjacent node (green color), and the nodes that share common boundaries with the neighboring node (white color)

Inspired by previous studies[14, 15], it was also assumed that the majority of nodes along the image boundaries belong to the real background and utilizes the "one-dimension rule" to deal with the situation in which that part of a salient object may touch the image boundaries. But different from using manifold ranking [14] or geodesic distance [15], the paper only adopted color contrast to measure the difference.

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

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

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, the neighborhood influence of node i was considered on the sparsely connected region R to obtain the final color contrast (CC):

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

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

Eq (8) and (11) were integrated to obtain the final refined conspicuity value of sp_i:

$$C_i' = C_i + CC_i \tag{12}$$

Spatial information has been widely used in many state-of-the-art saliency models [3, 7, 11, 12]. According to the phenomenon named as "center-bias", which originated from the humans' habit of placing the object in the middle of the image when capturing a photograph, the per-pixel conspicuity map was optimized to obtain the prior map. The value of pixel x in the prior map $p(s)$ is computed by:

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

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 the image, respectively. Fig. (**3h**), shows that the prior map highlights the real object while the previous Bayesian method in [19] is the part of redundant background salient (Shown in Fig. (**3d**)).

2.3. Enhanced Observation Likelihood

It is obvious that a more accurate prior map can greatly benefit the computation of observation likelihood. Different from the earlier study[19], which directly assumed all the pixels inside the convex hull to be foregrounded and these outer pixels to be in the background (see Fig. (**3c**)), this paper used the adaptive threshold strategy described in [24] to binarize the prior map (see Fig. (**3h**)). As shown in Fig. (**5a**), the enhanced estimation result is more consistent with the human visual observation result (Fig. (**5c**)) compared with the convex hull (Fig. (**3c**)).

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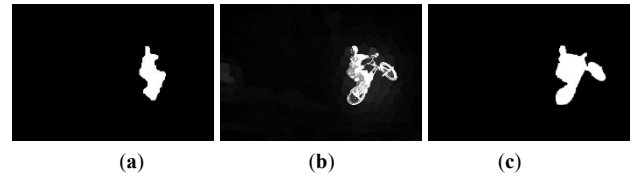
Fig. (5). (**a**) is the improved foreground and background estimation result after binarization, (**b**) is the saliency map of our method, (**c**) is the Ground Truth labeled by human

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

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

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

By substituting the $p(s)$ of (13), $p(x|s)$ of (14) and $p(x|b)$ of (15) into (1), the improved Bayesian saliency map was formed(Fig. (**5b**)).

3. EXPERIMENTS

This study evaluated 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 were 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]. In the parameters setting, each image was decomposed into 200 superpixels for efficiency. In (10), K=10 was set 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

Org.	Itti	AC	FT	RA	RC	CB	GS-SP

LR	SF	PCAS	GC	XL	BS	Our	GT
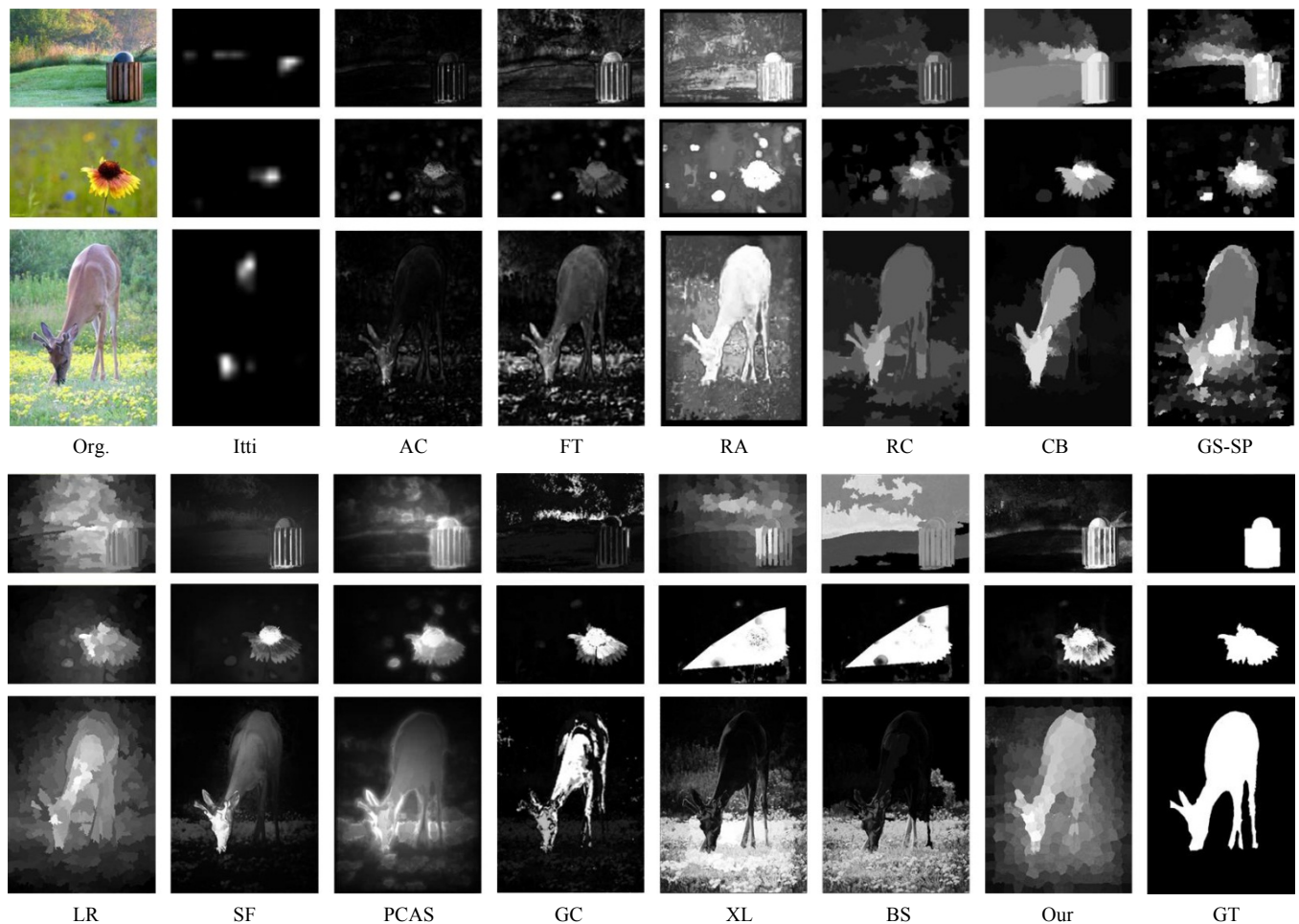
Fig. (6). The visual comparison of our saliency maps with 13 state-of-the-art methods.

methods is shown in Fig. (**6**) (Org. means the original image). It is obvious that the proposed method can highlight the real salient objects (the white pixels in GT) more accurately and uniformly than the alternatives. It is important to note that the improved Bayesian saliency model apparently outperformed 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 the 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 redundant background features into salient region improperly, inevitably leading to an inaccurate map (the first and second images of XL and BS in Fig. (**6**)). Even in the extreme case, the convex hull wrongly circles the part of background as the real object, which may greatly enhance the undesired background while ignoring the whole object in the saliency map (the third image of XL and BS in Fig. (**6**)). These shortcomings were not observed in the method employed in this study.

This study adopted 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 the detected salient pixels, while recall means the percentage of correct salient pixels with respect to the GT. As shown in Fig. (**7**), this method outperformed others when the recall value was under 0.85. When the recall was in the range [0.85, 1], the precision of this method was slightly lower than the GS-SP, but still higher than most of the compared methods. The highest precision value of our method was 0.95, superior to GS-SP (0.91). Compared to the three Bayesian saliency models: RA, XL and BS, the precision of this improved model was much higher.

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

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

Where β2=0.3 to assign higher precision according to [9-11]. This paper calculated all the F-measures on the thresholds ranging from 0 to 255 just like determining the P-R curves [20]. As shown in Fig. (**8**), this method has the highest F-measure 0.871 when T=183 (CB method [24] has the second highest

F-measure 0.866 when T=125) and it maintained the high level when T varied in a wide range [100, 225]. The F-measure of RA (0.727), XL (0.832) and BS (0.852) was
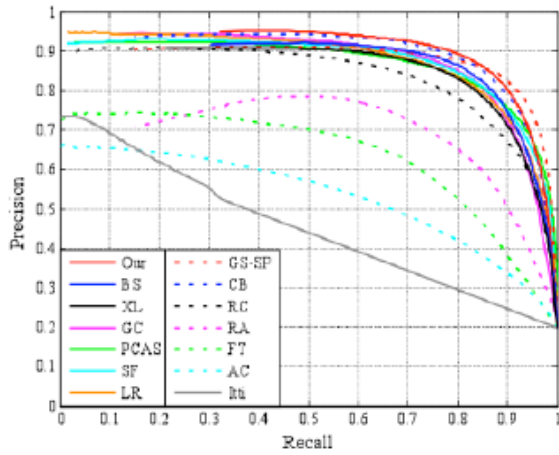
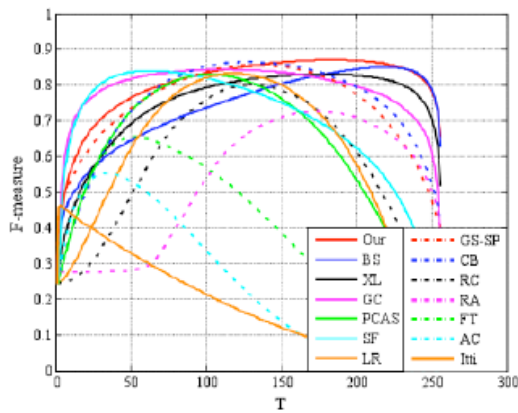**Fig. (7).** The P-R curves of different methods.

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

much lower than the proposed method , indicating that the proposed method indeed improved the Bayesian Framework.

This study also adopted the adaptive threshold strategy explained in [9] which is defined as twice of the mean salient value used to obtain the average precision, recall and F-measure (see Fig. (**9**)). It was concluded from the comparisons that: (1) The F-measure of the proposed method (0.836) outperformed 12 state-of-the-art methods except CB (0.861) which employed the global context and shape salient clues. (2) Compared to RA, XL and BS, the precision and F-measure of the proposed method was best, demonstrating the efficiency of the improvement.

The above P-R curve, F-measure and adaptive threshold measure do not consider the number of pixels that is accurately marked as the non-salient region (the black pixels in GT, as shown in Fig. (**6**)). To obtain more balanced comparison results, Mean Absolute Error (MAE) was introduced in [10]:

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

Where, W and H are the width and height of the saliency map. The MAE histograms are shown in Fig. (**10**). This improved method produced 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).
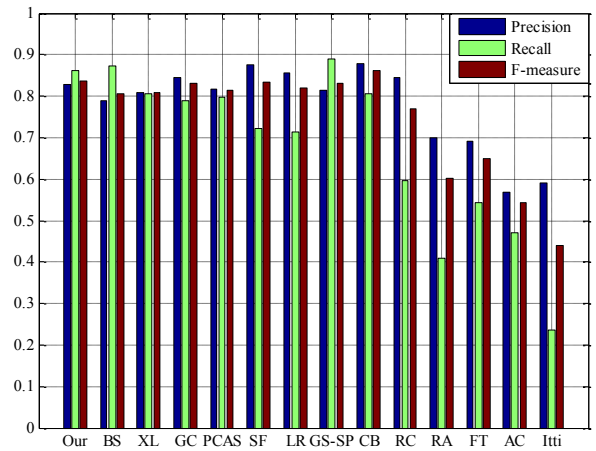

Fig. (9). The precision, recall and F-measure after adaptive threshold.
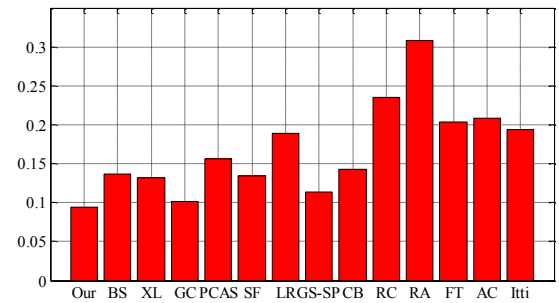

Fig. (10). The MAE histograms of different methods.

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 a study [26], due to the complexity of SOD, this paper selected   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 the proposed method. The results of all   four quantitative evaluation methods were not impressive.

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

Similar to the comparison results of P-R curves, as shown in Fig. (**11b**), the highest F-measure of the method (0.616) was only lower than the GS-SP (0.650), but still higher than others, especially RA,XL and BS.
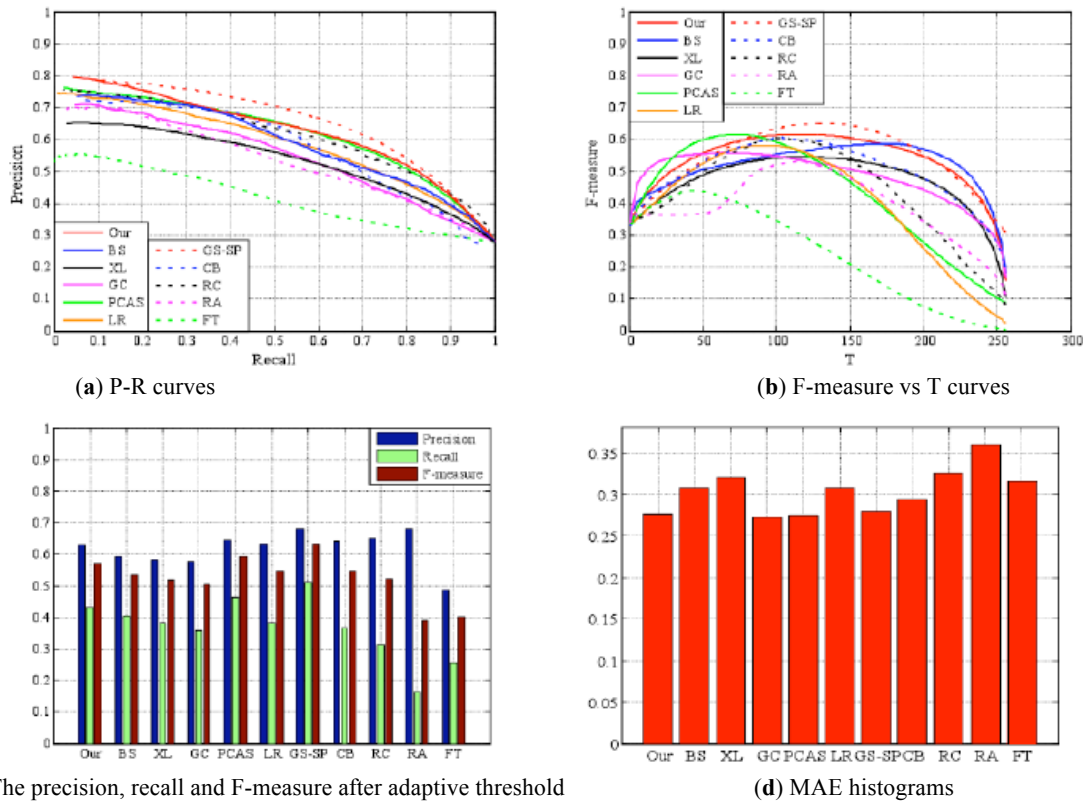
After adaptive threshold, the average F-measure of the method (0.569) was lower than PCAS (0.592) and GS-SP

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

(0.632), but higher than RA (0.39), XL (0.519) and BS (0.535) (See Fig. (**11c**)). 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. (**11d**), the MAE value of the proposed method was0.276, 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 the proposed method has better overall performance than most state-of-the-art methods, including three previous Bayesian saliency models, but still needs further improvement when confronting the difficult images with multiple objects.

Several representative saliency maps of different methods are shown in Fig. (**12**). Considering the large-scale object; the first image in Fig. (**12**)) as an example, the proposed method produced the best result, which is very similar

to GT. By considering the three previous Bayesian saliency models, it can be observed that RA only highlighted a small part of salient object; XL wrongly assigned part of the background with high salient value; and BS generated lower salient value in some important part of the real object. For the low contrast image and the image with two different objects (the second and third image in Fig. (**12**)), the proposed method can also produce excellent visualization results, verifying the efficiency of the improvements made.

3.3. Failure Cases

The proposed method is based on the Bayesian model. Although, great improvements were made to obtain more accurate prior map and observation likelihood in the Bayesian framework, but still there were some images which did not accord with the proposed principle. Some failure cases are shown in Fig. (**13**). When the color of real salient object was very close to the background, especially the existing multiple objects with similar color of the background, the method failed. In these cases, RA, XL and BS generated worse results.
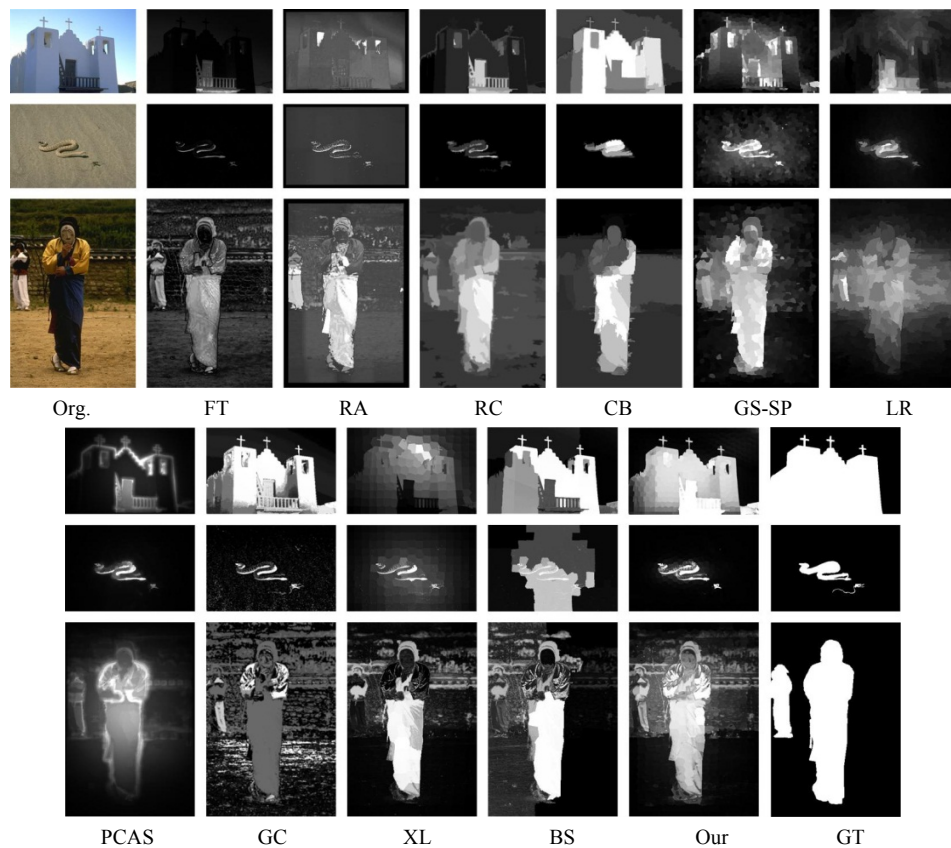
Org.   FT   RA   RC   CB   GS-SP   LR

PCAS   GC   XL   BS   Our   GT

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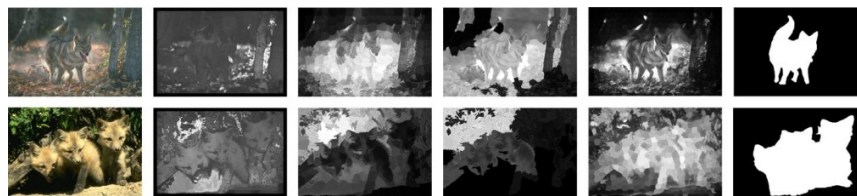

Fig. (13). Some failure saliency maps of our method and three previous Bayesian saliency models.

3.4. Running Time Comparison

Finally, this paper compared the average running time of these methods on MSRA-1000 and SOD datasets. Experiments were conducted using a computer with Inter(R) Core(TM)i5-2410M of 2.8GHz CPU and 8GB RAM. The proposed method was implemented by Matlab and the average running time was 3.32 s for an image, which is not very fast. However, considering the evaluation of quantitative performance, the proposed method still has advantage over these state-of-the-art methods Table **1**.

CONCLUSION

This paper proposed 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 more accurately provide the coarse location of potential object compared with the convex hull based Bayesian saliency model.The graph based color contrast measure greatly enhanced the prior map, which also benefitted the observation likelihood. After integration *via* Bayesian formula, the final saliency map was obtained with

better performance. The experimental results indicate that the proposed method outperformed all 13 sate-of-the-art methods in qualitative and quantitative evaluations, especially the methods based on the similar Bayesian framework−RA,XL and BS. In future work, we will investigate the discriminative salient features to more accurately estimate the location of potential object .

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

REFERENCES

[1]   A. Borji, and 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, and 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, and L. Itti, "Salient object detection: a benchmark", In: *Proc. European Conference on Computer Vision, Springer Press*, 2012, pp. 414-429.

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

[5]   L. Itti, C. Koch, and 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, and S. Süsstrunk, "Salient region detection and segmentation", In: *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, N. Zheng, X. Tang, and H.Y. Shum, "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, and 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, and S. Süsstrunk, "Frequency-tuned salient region detection", In: *Proc. IEEE International Conference on Computer Vision and Pattern Recognition,* IEEE Press, 2009, pp. 1597-1604.

[10]   F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection", In: *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, and C. Nigel, "Efficient salient region detection with soft iamge abstraction" ,In: *Proc. IEEE International Conference on Computer Vision,* IEEE Press, 2013, pp. 1529-1536.

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

[13]   Q. Yan, L. Xu, J. P. Shi, and J. Y. Jia, "Hierarchical saliency detection", In: *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, and M. H. Yang, "Saliency detection *via* graph-based manifold ranking", In:

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

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

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

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

[19]   Y. Xie, H. Lu, and 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, and S. F. Li, "Saliency detection based on integration of boundary and soft-segmentation", In: *Proc. IEEE International Conference on Image Processing,* IEEE Press, 2012 , pp. 1085-1088.

[21]   M. M. Cheng, Z. M. Zhang, W. Y. Lin, and P. Torr, "BING: Binarized normed gradients for objectness estimation at 300fps", In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* IEEE Press, 2014 , pp. 3286 -3293.

[22]   S. Hare, A. Saffari, and P. H. Torr, "Efficient online structured output learning for keypoint-based object tracking," In: *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, and 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, and S. Li, "Automatic salient object segmentation based on context and shape prior", In: *Proc. British Machine Vision Conference,* BMVA Press, 2011, pp. 110.1-110.12.

[25]   X. Shen, and Y. Wu, "A unified approach to salient object detection *via* low rank matrix recovery", In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* IEEE Press, 2012, pp. 853-860.

[26]   B. Jiang, L. Zhang, H. Lu, C .Yang, and M. Yang, "Saliency detection *via* absorbing markov chain," In: *Proc. IEEE International Conference on Computer Vision,* IEEE Press, 2013, pp. 1665-1672.

Proc. IEEE International Conference on Computer Vision and Pattern Recognition, IEEE Press, 2013, pp. 3166-3173.