Salient Object Detection with Complex Scene based on Cognitive Neuroscience

Chunbiao Zhu[1], Ge Li[1*], Wenmin Wang[1], Ronggang Wang[1]

[1]School of Electronic and Computer Engineering, Shenzhen Graduate School, Peking University
Shenzhen, China
e-mail: zhuchunbiao@pku.edu.cn; *corresponding author, geli@ece.pku.edu.cn

***Abstract*—Detecting salient objects with complex backgrounds is still a challenging problem. Under the background having similar colors with complex patterns of salient objects, existing methods' performance is not satisfied, especially for multiple salient objects detection. In this paper, we propose a framework based on cognitive neuroscience to tackle with these challenges. According to cognitive neuroscience, human visual system is sensitive to depth of field, conspicuous color, moving objects and central object of scene. In the proposed framework, we imitate these human visual characteristics with following approaches: (1) using depth to represent the depth of field in the real world; (2) using luminance which imitates the light changing to represent the relative motions among objects; (3) using the center-bias to enhance object around the center. Experimental results on two challenging RGB-D datasets demonstrate that our method is superior to the existing methods in terms of effectiveness.**

***Index Terms*—Salient object detection, Depth information, Center-bias, Luminance map**

I. Introduction

SALIENCY detection is a process of getting the visual attention region precisely. The attention is the behavioral and cognitive process of selectively concentrating on one aspect of the environment while ignoring other things.

In early time, it is defined as locating and tracking the visual attention region, thus most of works focus on predicting the eye-fixations [1].

Recently, it is extended to locate and refine the salient object, known as salient object detection. Many applications, which use salient object detection as a useful tool in the preprocessing, have been implemented, such as image retrieval [2], object recognition [3], object segmentation [4], image retargeting [5], image compression [6], and so on.

The works on salient detection mainly use heuristic saliency features and discriminative saliency features. Various methods based on heuristic saliency features have been proposed, such as pixel-based or patch-based contrast [7], region-based contrast [8], pseudo-background [9], and similar images [10], etc. There are some proposed methods using discriminative saliency features, such as multi-scale contrast [11], center-surround contrast [12], and color spatial compactness [13], etc. Besides, some other methods use image over-segmentation [14], outlier [15] or depth cue [16-19]. However, the above methods are not robust enough in challenging cases, where the background has similar colors with salient objects, and especially with multiple salient objects in the scene.

To improve the performance, Zhai et al [13] introduce image histograms to calculate the color saliency. Cheng et al [8] extend the histogram to 3D color space. These methods can find pixels/regions with colors much different from the dominant one. However, they do not consider spatial locations.

Besides, there are many other attempts in saliency detection by taking depth cue into consideration. Niu et al. [16] define saliency using depth cue computed from stereo images. Their results show that stereo saliency is a useful consideration compare to previous visual saliency analysis. Cheng et al [17], Tang et al. [18] and Peng et al [19] combine color and depth features to detect the single salient object in a scene. These methods are not robust in challenging cases, where multiple objects exist in one scene.

To overcome those challenges, we propose a detection framework based on cognitive neuroscience. Our approach imitates human visual system. Using image depth information to represent the depth of real field. Using space-variant luminance change to represent relative motions among objects. Using center-bias information to represent the fovea with which human visual systems locate object. Using color information to represent the selective characteristics of human eyes with different wavelengths. At last, we fuse all the information to get the final saliency map.

The proposed framework is consisted of three steps: 1) the preprocessing stage, including center-bias extraction and image luminance enhancement; 2) the preliminary stage, which utilizes both color features and depth features to obtain preliminary detection map; 3) the fusion stage, which fuses depth and center-bias information with preliminary detections to produce the final detection result. The saliency quality is evaluated by precision rate, recall rate, and mean absolute error. Based on two RGB-D datasets, experiment results show that our detection method is in accord with human subjective vision perception, which achieve high precision and recall rate, and less mean absolute errors compared with other methods.

The contributions of the paper include three folds: 1) we

This work was supported by the grant of National Science Foundation of China (No.U1611461), the grant of Science and Technology Planning Project of Guangdong Province, China (No.2014B090910001) and the grant of Shenzhen Peacock Plan (No.20130408-183003656).

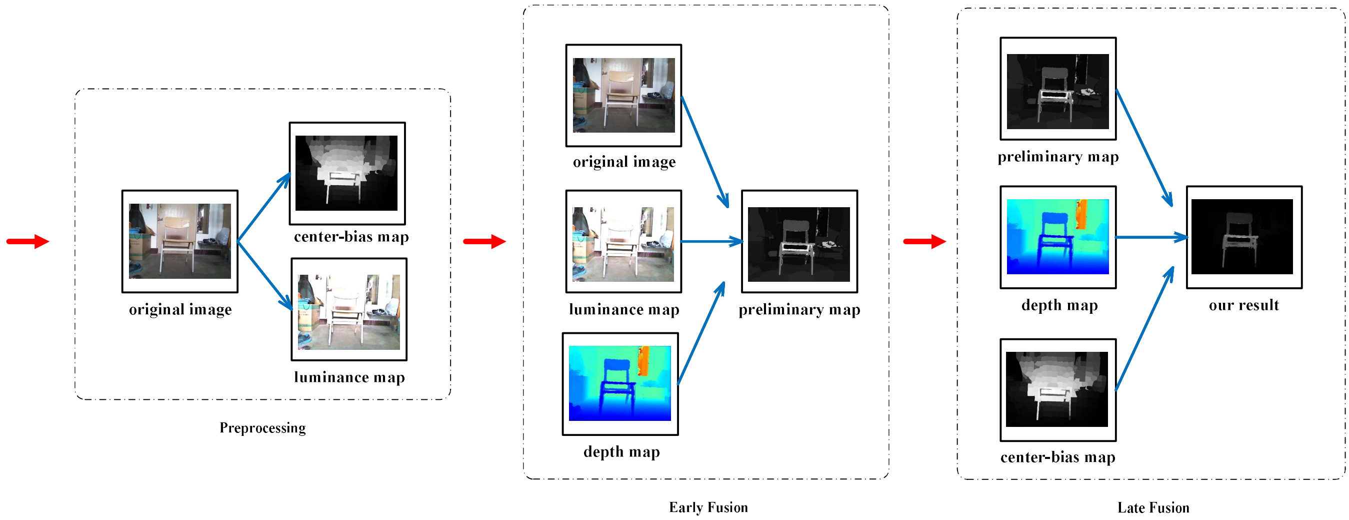


Fig. 1. The proposed framework of the salient detection in three steps: preprocessing, early fusion and late fusion. To make the depth map better represented, we render it with different colors.

apply cognitive neuroscience to deal with salient object detection which to make the saliency detection results accord with human vision perception; 2) we utilize two-stages fusion approach to obtain salient regions. At each stage, multiple features are extracted and used to exploit their complementary advantages; 3) we imitate the relative motion among objects via changing illuminance produced by image enhancement algorithm to enhance the contrast between salient objects and backgrounds.

II. Our Approach

Our method starts from preprocessing, by which we obtain center-bias saliency map and the image space-variant luminance map. Then, we compute saliency cues of original image, luminance map and depth map, by which the preliminary saliency map is produced. Finally, we fuse preliminary saliency map and center-bias saliency map to get final saliency detection result. The framework of the proposed approach is illustrated in Fig. 1.

A. Preprocessing of image

According to the results of cognitive neuroscience [20], human eye uses central fovea to locate object and make them clearly visible. Therefore, most of images taken by cameras always locate salient object around the center. Aiming to get center-bias saliency map, we use BSCA algorithm [21]. It construct global color distinction and spatial distance matrix based on clustered boundary seeds and integrate them into a background-based map. Thus it can improve the center-bias, erasing the image edge effect. As shown in the preprocessing stage of Fig. 1, the center-bias saliency map can remove the surroundings of the image and reserve most of salient regions. We denote this center-bias saliency map as S_{pc}.

A moving object can attract the human visual attention consistently. How to turn the static image into a motion image? We propose a novel method that uses the SVLM algorithm [22]. By using this algorithm, we imitate the light changing in the real world, from which the same scene at different time can be obtained. We use the luminance map to represent light changing which can reflect relative motion among objects. Here, for computation efficiency, we just take one luminance map. We denote the luminance map computed by the algorithm as I_s, and we define the final luminance map as I_l in equation (1):

$$I_1 = I + I_s \tag{1}$$

where I represents the original image, + stands for pixel to pixel gray summation, both images are normalized before the summation.

As shown in Fig. 1, the luminance map represent the counterpart of the original image under a different light condition, from which the contrast between salient object and background is enhanced.

B. Early fusion

In this stage, we extract color and depth features, and then mixes center-bias weight to enhance the saliency map. We denote the depth images as I_d, which are obtained from datasets [17] [19].

Inspired by the work in [8] [14], we calculate the saliency value via the method based on spatially weighted region contrast.

Firstly, the image is segmented into K regions based on color and depth cues via the K-means algorithm.

The color saliency of region k, where $k \in [1, K]$, is defined as:

$$S_c(r_k) = \sum_{i=1, i\neq k}^{K} P_i W_d(r_k)\, D_c(r_k, r_i) \tag{2}$$

where r_k and r_i represent regions k and i respectively, $D_c(r_k,r_i)$ is the Euclidean distance between region k and region i in L*a*b color space, P_i represents the area ratio of region r_i compared with the whole image, $W_d(r_k)$ is the spatial weighting term of the region k, defined as:

$$W_d(r_k) = e^{-\frac{D(r_k, r_i)}{\sigma^2}} \tag{3}$$

where $D(r_k,r_i)$ is the Euclidean distance between the centers of region k and i, σ is the parameter controlling the strength of W_d, we take $\sigma^2 = 0.4$ which has the best contribution to the results.

Similar to color saliency, we define the depth saliency of a map as:

$$S_{\mathrm{d}}(r_k) = \sum_{i=1, i\neq k}^{K} P_i W_d(r_k)\ D_d(r_k, r_i) \tag{4}$$

where $D_d(r_k, r_i)$ is the Euclidean distance between region k and region i in depth space.

We calculate color saliency images for original image I and luminance image I_l respectively by using formula (2), which are denoted as $S_{oc}(r_k)$ and $S_{lc}(r_k)$. Then, we get the final color saliency value via equation (5):

$$S_{\mathrm{fc}}(r_k) = S_{lc}(r_k) + S_{oc}(r_k) \tag{5}$$

by implementing such processing, the co-salient regions within the two images are enhanced.

TABLE I

THE PSEUDOCODE OF OUR FRAMEWORK

Algorithm Framework of our saliency detection method

Input: original map I, depth map I_d, center-bias map S_{pc}, luminance map I_l
Output: saliency value S
for each region k and i ∈ [1, K] **do**
compute the color saliency value of the original map $S_{oc}(r_k) = P_i W_d(r_k) D_c(r_k, r_i)$
get the color saliency value of the luminance map $S_{lc}(r_k) = P_i W_d(r_k) D_c(r_k, r_i)$
obtain the final color saliency value $S_{fc}(r_k) = S_{oc}(r_k) + S_{lc}(r_k)$
compute the depth saliency value $S_d(r_k) = P_i W_d(r_k) D_d(r_k, r_i)$
calculate the center-bias and depth weights $W_{cd}(r_k) = DW(d_k) \times G(\|P_k - P_o\|)/N_k$
the preliminary saliency value is $S_p(r_k) = G(\{S_{fc}(r_k) + S_d(r_k)\} \times W_{cd}(r_k))$
the final saliency value can be caculated via $S(r_k) = S_p(r_k) \times (\neg I_d(r_k) \times S_{pc}(r_k)$
final
return the saliency value of the image S

For most images, the salient object always locate at the center of the image or close to the camera. Therefore, we add the center-bias and depth weights for both color and depth images. The weights of the region k are defined as:

$$W_{\mathrm{cd}}(\mathrm{r}_k) = \frac{G\left(\|\mathrm{P}_{\mathrm{k}} - \mathrm{P}_{\mathrm{o}}\|\right)}{N_k} DW(d_k) \tag{6}$$

where G(.) represents the Gaussian normalization, ||.|| is two regions' position Euclidean distance, P_k is the position of the region k, P_o is the center position of this map, N_k is the pixel number of region k, $DW(d_k)$ is the depth weight, which is defined as:

$$DW(d_k) = (\max\{d\} - d_k)^{\beta} \tag{7}$$

where max {d} represents the maximum depth in the image, and d_k is the depth value of region k, β is a fixed value for a depth map, defined as:

$$\beta = \frac{1}{\max\{d\} - \min\{d\}} \tag{8}$$

where min {d} represents the minimum depth of the image.

Then, the preliminary saliency map is defined as:

$$S_{\mathrm{p}}(r_k) = G\left(\{S_{\mathrm{f}c}(r_k) + S_d(r_k)\} \times W_{cd}(r_k)\right) \tag{9}$$

C. Late fusion

So far, we have the preliminary saliency map. To refine the salient detection results, we optimize the preliminary saliency map with the help of the center-bias and depth maps. The final saliency map is defined as following:

$$S(r_k) = S_{\mathrm{p}}(r_k) \times (\neg I_{\mathrm{d}}(\mathrm{r}_k)) \times S_{pc}(\mathrm{r}_k) \tag{10}$$

where S_{pc} is center-bias saliency map denoted in preprocessing, I_d is the depth image, and ¬ is the negation operation which enhance the saliency degree of front regions, and $S_p(r_k)$ is defined in equation (9).

The pseudocode of our framework is shown in Table I.

III. EXPERIMENTS AND EVALUATION

A. RGB-D Dataset 1

In our experiments, the RGB-D saliency detection dataset [17] are applied. This dataset has 135 indoor images taken by Kinect with the resolution 640×480. This dataset has the complex backgrounds and irregular shapes of salient objects.

For performance evaluation, precision-recall curve and MAE (Mean Absolute Error) are utilized as common in literature of saliency detection. The MAE is formulated as:

$$MAE = \frac{\sum_{i=1}^{N} \left|GT_{\mathrm{i}} - S_i\right|}{N} \tag{11}$$

where N is the number of the images, GT_i is area of the ground truth on image i, S is the area of detection result.

We compare our method with existing methods, such as BSCA [21], HS [23], DES [17], LPS [24], FT [7], and SIM [25]. We use the authors' original codes provided by the author to implement corresponding experiments.

The quantitative evaluation results are shown in Fig. 2.

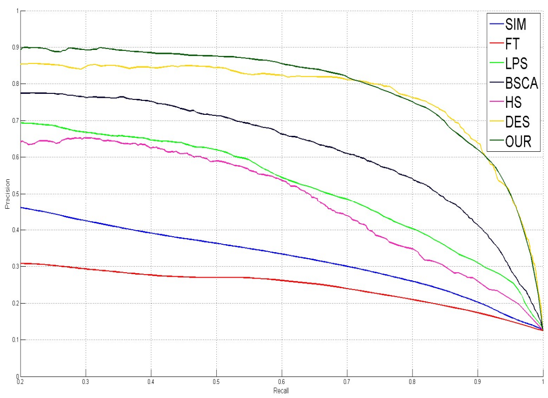


Fig. 2. Comparison results on the RGB-D dataset 1 [17].

From the precision-recall curve, we can see that our method achieves better precision and recall rate.

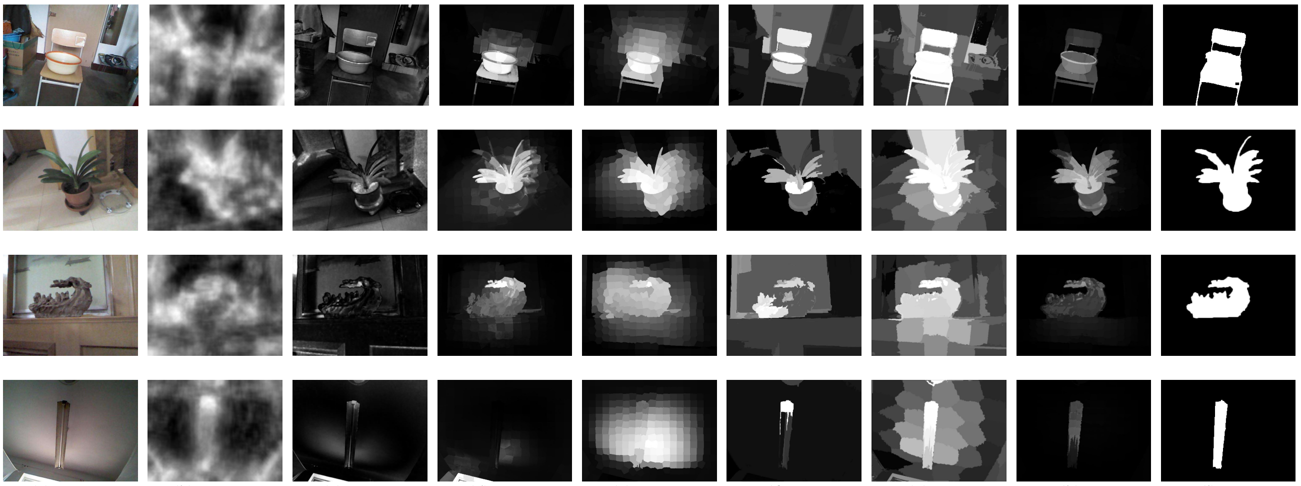
MAE results are shown in Table II, where the lower value, the better performance. Compared with the MAE value, it can be observed that our method obtains more precise salient regions than that of other approaches.

TABLE II

MAE RESULTS ON RGBD DATASET 1 [17]

Method	MAE
SIM	0.3740
FT	0.2150
LPS	0.1406
BSCA	0.1851
HS	0.1849
DES	0.3079
OUR	0.1065

The smallest value, which is the best, is marked with red color.



(a) Image (b) SIM (c) FT (d) LPS (e) BSCA (f) HS (g) DES (h) OUR (i) GT

Fig. 3. Visual comparison of saliency maps produced by various methods on RGB-D dataset 1 [17]. (a) Original Image. (b) SIM [25]. (c) FT [7]. (d) LPS [24]. (e) BSCA [21]. (f) HS [23]. (g) DES [17]. (h) Our Method. (i) Ground Truth.

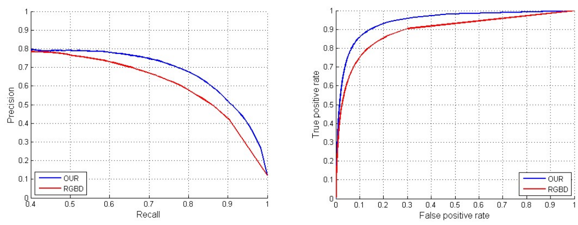
In Fig. 3, some real examples of the detection results are given with explicit images for a subjective comparison. It clear shows that our method detects the saliency regions more precisely, compared with other approaches.

B. RGB-D Dataset 2

In our experiments, we also use another challenging RGB-D saliency detection dataset [19], which contains 400 kinds of common objects captured in 11 types of scenes under different illumination conditions. The indoor scenes include offices, apartments, supermarkets, museums, etc., while the outdoor locations cover parks, campuses, streets, etc. Most images contain single salient object, while the other images include multiple objects. It has totally 1000 image with two different resolutions of 640×480 and 480×640. Compared with RGB-D dataset 1, RGB-D dataset 2 is more challenging.

Our method is also evaluated by the measures including precision-recall curve and MAE (Mean Absolute Error). Besides, we utilize ROC (Receiver Operating Characteristic) to evaluate the robustness of the detection results.

We compare our method only with the RGBD algorithm [19]. Because in the DIWC approach proposed in [18], the author states that theirs method has some limitations on RGB-D dataset 2, and they verify that the RGBD algorithm has better detection results than many other methods including theirs. We use the codes provided by the author to reproduce their experiments.



(a) PR (b) ROC

Fig. 4. Comparison results on the RGBD dataset 2. (a) the PR curve and (b) the ROC curve.

The quantitative evaluation results are shown in Fig. 4. Observed from the precision-recall curve in Fig. 4(a), we can see that our method achieves high precision and recall rate simultaneously, which demonstrates that the proposed approach is superior to RGBD algorithm.

The ROC evaluation results are shown in Fig. 4(b), which also proves that our method outperforms than RGBD algorithm.

The evaluation results of MAE are shown in Table III. It can observed that our method can obtain more precise saliency detection results than RGBD.

Similarly, the visual comparisons are given in fig. 5, which vividly demonstrate the advantages of our approach. We can see that our method can detect both single salient object and multiple salient objects more precisely.

IV. Conclusions And Future Work

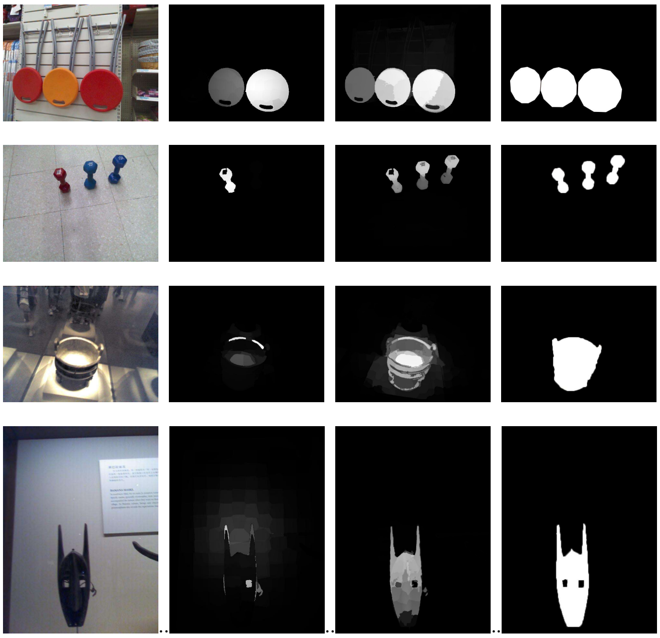
TABLE III
MAE RESULTS ON RGBD DATASET 2 [19]

Method	MAE
RGBD	0.1048
OUR	0.1007

The smallest value, which is the best, is marked with red color.

We proposed a novel method of detecting salient objects under complex scene based on cognitive neuroscience. Our approach included three steps: firstly, we referred to cognitive neuroscience to extract the center-bias saliency map and luminance map from the original image, which imitated human visual characteristic; then, we fused depth and color cues with spatial weights to obtain preliminary saliency map; finally, for getting robust detection results, we fused the preliminary map and depth map and center-bias map. To encourage future works, we make the source code open, as well as related materials. More testing results can be found on our project website: https://github.com/ChunbiaoZhu/BCNS.

In the future, we will further explore human visual



(a) Original Image (b) RGBD (c) OUR (d) GT

Fig. 5. Visual comparison of saliency maps produced by various methods on RGB-D dataset 2. (a) Original Image. (b) RGBD [19]. (c) Our Method. (d) Ground Truth.

characteristic and mechanism of human consciousness, based on which we may extract some new features to enhance the saliency map.

REFERENCES

[1] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1998, 20(11):1254-1259.

[2] Cheng M M, Mitra N J, Huang X, et al. SalientShape: group saliency in image collections[J]. The Visual Computer, 2014, 30(4):443-453.

[3] Alexe B, Deselaers T, Ferrari V. Measuring the Objectness of Image Windows[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, 34(11):2189-202.

[4] Donoser M, Urschler M, Hirzer M, et al. Saliency driven total variation segmentation[C]// IEEE, International Conference on Computer Vision. IEEE, 2009:817-824.

[5] Sun J, Ling H. Scale and object aware image retargeting for thumbnail browsing[C]// IEEE International Conference on Computer Vision. 2011:1511-1518.

[6] Itti L. Automatic foveation for video compression using a neurobiological model of visual attention[J]. IEEE Transactions on Image Processing, 2004, 13(10):1304-18.

[7] Achanta R, Hemami S, Estrada F, et al. Frequency-tuned salient region detection[C]// IEEE International Conference on Computer Vision and Pattern Recognition. 2009:1597-1604.

[8] Cheng M M, Zhang G X, Mitra N J, et al. Global contrast based salient region detection[C]// Computer Vision and Pattern Recognition. IEEE, 2015:409-416.

[9] Wei Y, Wen F, Zhu W, et al. Geodesic Saliency Using Background Priors[C]// European Conference on Computer Vision. 2012:29-42.

[10] Marchesotti L, Cifarelli C, Csurka G. A framework for visual saliency detection with applications to image thumbnailing[C]// International Conference on Computer Vision IEEE International Conference on Computer Vision. 2009:2232-2239.

[11] Liu T, Sun J, Zheng N N, et al. Learning to Detect A Salient Object[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007:1-8.

[12] Jiang H, Wang J, Yuan Z, et al. Salient Object Detection: A Discriminative Regional Feature Integration Approach[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2013:2083-2090.

[13] Mehrani P. Saliency Segmentation based on Learning and Graph Cut Refinement.[J]. 2010.

[14] Cheng M M, Warrell J, Lin W Y, et al. Efficient Salient Region Detection with Soft Image Abstraction[C]// textbfIEEE ICCV. 2013:1529-1536.

[15] Shen X, Wu Y. A unified approach to salient object detection via low rank matrix recovery[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2012:853-860.

[16] Niu Y, Geng Y, Li X, et al. Leveraging stereopsis for saliency analysis[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2012:454-461.

[17] Cheng Y, Fu H, Wei X, et al. Depth Enhanced Saliency Detection Method[J]. European Journal of Histochemistry Ejh, 2014, 55(1):301-308.

[18] Tang Y, Tong R, Tang M, et al. Depth incorporating with color improves salient object detection[J]. The Visual Computer, 2016, 32(1):111-121.

[19] Peng H, Li B, Xiong W, et al. RGBD Salient Object Detection: A Benchmark and Algorithms[J]. Lecture Notes in Computer Science, 2014, 8691:92-109.

[20] Baars B J, Gage N M. - Cognition, Brain, and Consciousness (Second Edition)[J]. South African Journal of Occupational Therapy, 2008, 41(1):591–616.

[21] Qin Y, Lu H, Xu Y, et al. Saliency detection via Cellular Automata[J]. 2015:110-119.

[22] Lee S, Kwon H, Han H, et al. A Space-Variant Luminance Map based Color Image Enhancement[J]. IEEE Transactions on Consumer Electronics, 2010, 56(4):2636-2643.

[23] Shi J, Yan Q, Xu L, et al. Hierarchical Image Saliency Detection on Extended CSSD.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 38(4):1.

[24] Li H, Lu H, Lin Z, et al. Inner and Inter Label Propagation: Salient Object Detection in the Wild[J]. Image Processing IEEE Transactions on, 2015, 24(10):3176-3186.

[25] Murray N, Vanrell M, Otazu X, et al. Saliency estimation using a non-parametric low-level vision model[J]. 2011, 42(7):433-440.