A Multilayer Backpropagation Saliency Detection Algorithm Based on Depth Mining

Chunbiao Zhu[1], Ge Li[1](✉), Xiaoqiang Guo[2], Wenmin Wang[1] and Ronggang Wang[1]

[1] School of Electronic and Computer Engineering, Shenzhen Graduate School, Peking University Shenzhen, China
[2] Academy of Broadcasting Science, SAPPRFT Beijing, China
zhuchunbiao@pku.edu.cn, ✉geli@ece.pku.edu.cn

Abstract. Saliency detection is an active topic in multimedia field. Several algorithms have been proposed in this field. Most previous works on saliency detection focus on 2D images. However, for some complex situations which contain multiple objects or complex background, they are not robust and their performances are not satisfied. Recently, 3D visual information supplies a powerful cue for saliency detection. In this paper, we propose a multilayer backpropagation saliency detection algorithm based on depth mining by which we exploit depth cue from four different layers of images. The evaluation of the proposed algorithm on two challenging datasets shows that our algorithm outperforms state-of-the-art.

Keywords: Saliency Detection, Depth Cue, Depth Mining, Multilayer, Backpropagation.

1 Introduction

Salient object detection is a process of getting a visual attention region precisely from an image. The attention is the behavioral and cognitive process of selectively concentrating on one aspect within the environment while ignoring other things.

Early work on computing saliency aims to locate the visual attention region. Recently the field has extended to locate and refine the salient regions and objects. Many saliency detection algorithms have been used as a useful tool in the pre-processing, such as image retrieval [1], object recognition [2], object segmentation [3], compression [4], image retargeting [5], etc.

In general, saliency detection algorithm mainly use top-down or bottom-up approaches. Top-down approaches are task-driven and need supervised learning. While bottom-up approaches usually use low-level cues, such as color features, distance features, depth features and heuristic saliency features. One of the most used heuristic saliency feature [6-10] is contrast, such as pixel-based or patch-based contrast, region-based contrast, multi-scale contrast, center-surround contrast, color spatial compactness, etc. Although those methods have their own advantages, they are not robust to

specific situations which lead to inaccuracy of results on challenging salient object detection datasets.

To deal with the challenging scenarios, some algorithms [11-15] adopt depth cue. In [11], Zhu et al. propose a framework based on cognitive neuroscience, and use depth cue to represent the depth of real field. In [12], Cheng et al. compute salient stimuli in both color and depth spaces. In [13], Peng et al. provide a simple fusion framework that combines existing RGB-produced saliency with new depth-induced saliency. In [14], Ju et al. propose a saliency method that works on depth images based on anisotropic center-surround difference. In [15], Guo et al. propose a salient object detection method for RGB-D images based on evolution strategy. Their results show that stereo saliency is a useful consideration compare to previous visual saliency analysis. All of them demonstrate the effectivity of depth cue in improvement of salient object detection.

Although, those approaches can enhance salient object region. It is very difficult to produce good results when a salient object has low depth contrast compared to the background. The behind reason is that only partial depth cue is applied. In this paper, we propose a multilayer backpropagation algorithm based on depth mining to improve the performance.

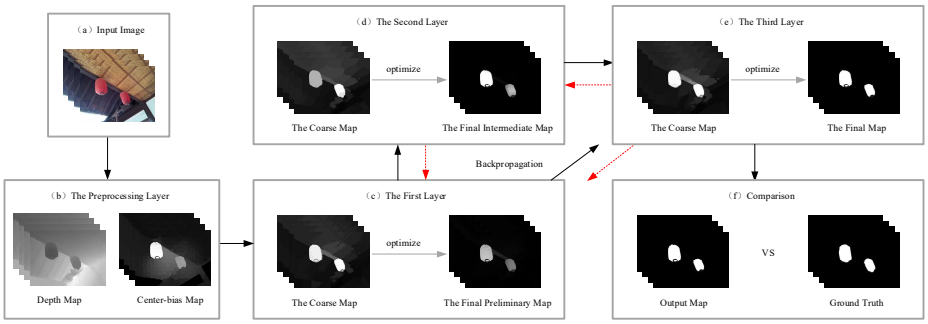

Fig. 1. The framework of the proposed algorithm.

2 Proposed Algorithm

The proposed algorithm is a multilayer backpropagation method based on depth mining of an image. In the preprocessing layer, we obtain the center-bias saliency map and depth map. In the first layer, we use original depth cue and other cues to calculate preliminary saliency value. In the second layer, we apply processed depth cue and other cues to compute intermediate saliency value. In the third layer, we employ reprocessed depth cue and other cues to get final saliency value. The framework of the proposed algorithm is illustrated in Fig. 1.

2.1 The Preprocessing Layer

In the preprocessing layer, we imitate the human perception mechanism to obtain center-bias saliency map and depth map.

Center-bias Saliency Map. Inspired by cognitive neuroscience, human eyes use 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 [6]. It constructs 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. 2 (c), 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 $\mathrm{C_b}$.

Depth Map. Similarly, biology prompting shows that people perceive the distance and the depth of the object mainly relies on the clues provided by two eyes, and we call it binocular parallax. Therefore, the depth cue can imitate the depth of real field. The depth map used in the experimental datasets is taken by Kinect device. And we denote the depth map as $\mathrm{I_d}$.

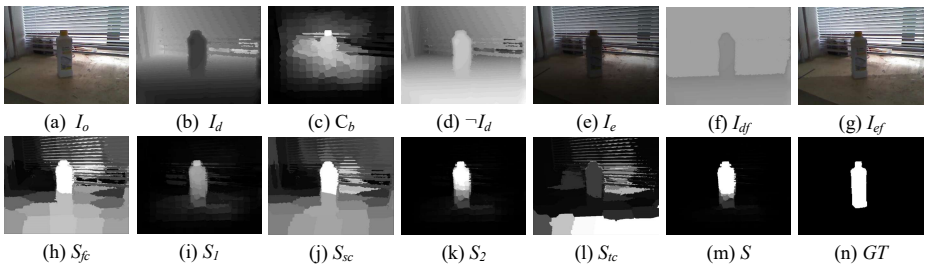

Fig. 2. The visual process of the proposed algorithm.

2.2 The First Layer

In the first layer, we extract color and depth features from the original image I_o and the depth map I_d, respectively.

First, the image I_o is segmented into K regions based on color via the K-means algorithm. Define:

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

where $S_c(r_k)$ is the color saliency of region k, $k \in [1, K]$, 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, $W_s(r_k)$ is the spatial weighted term of the region k, set as:

$$W_s(r_k) = \mathrm{e}^{-\frac{D_o(r_k, r_i)}{\sigma^2}}\,, \tag{2}$$

where $D_o(r_k, r_i)$ is the Euclidean distance between the centers of region k and i, σ is the parameter controlling the strength of $W_s(r_k)$.

Similar to color saliency, we define:

$$S_d(r_k) = \sum_{i=1,i\neq k}^{K} P_i W_s(r_k) D_d(r_k, r_i) \,, \tag{3}$$

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

In most cases, a salient object always locate at the center of an image or close to a camera. Therefore, we add the weight considering both center-bias and depth for both color and depth images. The weight of the region k is set as:

$$S_s(r_k) = \frac{G(\|P_k - P_o\|)}{N_k} W_d(d_k) \,, \tag{4}$$

where G(·)represents the Gaussian normalization, $\|\cdot\|$ is Euclidean distance, P_k is the position of the region k, P_o is the center position of this map, N_k is the number of pixels in region k, W_d is the depth weight, which is set as:

$$W_d = (max\{d\} - d_k)^{\mu} \,, \tag{5}$$

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

$$\mu = \frac{1}{max\{d\} - min\{d\}} \,, \tag{6}$$

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

Second, the coarse saliency value of the region k is calculated as:

$$S_{fc}(r_k) = G\big(S_c(r_k) S_s(r_k) + S_d(r_k) S_s(r_k)\big) \,, \tag{7}$$

Third, to refine the salient detection results, we optimize the coarse saliency map with the help of the center-bias and depth maps. The preliminary saliency map is calculated as following:

$$S_1(r_k) = S_{fc}(r_k) \neg I_d(r_k) \mathrm{C_b}(r_k) \,, \tag{8}$$

where ¬ is the negation operation which can enhance the saliency degree of front regions as shown in Fig. 2(d), because the foreground object has low depth value in depth map while the background object has high depth value.

2.3 The Second Layer

In the second layer, we exploit depth map further, we allocate the color values to the depth map according to different depth values, in this way, we can polarize the color attribute between foreground and background.

First, we set:

$$I_e\langle R|G|B\rangle = I_o\langle R|G|B\rangle \times I_d \,, \tag{9}$$

where I_e represents the extended color depth map. $\langle R|G|B\rangle$ represents processing of three RGB channels, respectively.

The extended color depth map is displayed in Fig. 2(e), from which the salient objects' edges are prominent.

Second, we use extended color depth map I_e to replace I_o. Then, we calculate intermediate coarse saliency value S_{sc} via the first stage' Eq. (1)-(6). We get:

$$S_{sc}(r_k) = G\big(S_c(r_k)S_s(r_k) + S_d(r_k)S_s(r_k)\big)\,, \tag{10}$$

where $S_{sc}(r_k)$ is the intermediate saliency value.

Third, to refine coarse saliency value, we apply the backpropagation to enhance the intermediate saliency value by mixing the result of the first layer. And we define our intermediate saliency value as:

$$S_2(r_k) = S_1^2(r_k) + S_1(r_k)\left(\mathbf{1} - \boldsymbol{e}^{-\mathrm{S}_{sc}{}^2(r_k)\,\neg I_d(r_k)}\right), \tag{11}$$

2.4 The Third Layer

In the third layer, we find that background noises can reduced by filtering the depth map, so, we exploit the depth map again.

First, we reprocess the depth cue by filtering the depth map via the following formula:

$$I_{df} = \begin{cases} I_d, & \mathrm{d} \le \beta \times \max\{\mathrm{d}\} \\ 0, & \mathrm{d} > \beta \times \max\{\mathrm{d}\} \end{cases}, \tag{12}$$

where I_{df} represents the filtered depth map. In general, salient objects always have the small depth value compared to background, thus, by Eq. (12), we can filter out the background noises. β is the parameter which controls the strength of I_{df}.

Second, we extend the filtered depth map to the color images via the Eq. (9). We denote the reprocessed depth map as I_{ef}.

We use filtered depth map I_{ef} to replace I_o. Then, we calculate third coarse saliency map S_{tc} via the first stage' Eq. (1)-(6), denoted as:

$$S_{tc}(r_k) = G\big(S_c(r_k)S_s(r_k) + S_d(r_k)S_s(r_k)\big)\,, \tag{13}$$

Fourth, to refine $S_{tc}(r_k)$, we apply the backpropagation of $S_1(r_k)$ and $S_2(r_k)$ as following:

$$S(r_k) = S_2(r_k)\big(S_2(r_k) + S_{tc}(r_k)\big)\Big(S_{tc}(r_k) + \mathbf{1} - \boldsymbol{e}^{-\,S_{tc}^2(r_k)S_1(r_k)}\Big)\,. \tag{14}$$

From the Fig. 2, we can see the visual results of the proposed algorithm. The main steps of the proposed salient object detection algorithm are summarized in Algorithm 1.

Algorithm 1

Input: original map I_o, depth map I_d

Output: final saliency value $\mathrm{S}(r_k)$

1: **for** each region $k = [1, K]$ **do**

2: compute color saliency value $S_c(r_k)$and depth saliency value $S_d(r_k)$

3: calculate the center-bias and depth weights $W_{cd}(r_k)$

4: get the preliminary saliency value $S_1(r_k)$

5: **end for**

6: **repeat** step 1-5 for the extended depth maps I_e , then calculate$S_2(r_k)$

7: **repeat** step 1-5 for the filtered depth maps I_{ef} , then calculate $S_{tc}(r_k)$

8: calculate $S(r_k)$ by applying the backpropagation of $S_1(r_k)$ and $S_2(r_k)$

9: **return** the final saliency value $\mathrm{S}(r_k)$

3 Experimental Evaluation

3.1 Datasets and Evaluation Indicators

Datasets. We evaluate the proposed saliency detection algorithm on two RGBD standard datasets: RGBD1* [12] and RGBD2* [13]. RGBD1* has 135 indoor images taken by Kinect with the resolution of 640×480. This dataset has complex backgrounds and irregular shapes of salient objects. RGBD2* contains 1000 images with two different resolutions of both 640×480 and 480×640, respectively.

Evaluation Indicators. Experimental evaluations are based on standard measurements including precision-recall curve, ROC curve, MAE (Mean Absolute Error), F-measure. The MAE is formulated as:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N} |GT_i - S_i|}{N}. \tag{15}$$

And the F-measure is formulated as:

$$F\text{-}measure = \frac{2\times\mathrm{Precision}\times\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Reca}}. \tag{16}$$

3.2 Ablation Study

We first validate the effectiveness of each layer in our method: the first layer result, the second layer result and the third layer result. Table. I shows the MAE and F-measure validation results on two datasets. We can clearly see the accumulated processing gains after each layer, and the final saliency result shows a good performance. After all, it proves that each layer in our algorithm is effective for generating the final saliency map.

Table 1. Validation results on two datasets.

	RGBD1* Dataset			RGBD2* Dataset		
Layers	S_1	S_2	S	S_1	S_2	S
MAE Values	0.1065	0.0880	0.0781	0.1043	0.0900	0.0852
F-measure Values	0.5357	0.6881	0.7230	0.5452	0.7025	0.7190

3.3 Comparison

To illustrate the effectiveness of our algorithm, we compare our proposed methods with DES14 [12], RSOD14 [13], BSCA15 [6], LPS15 [10], ACSD15 [14], HS16 [9] and SE16 [11]. We use the codes provided by the authors to reproduce their experiments. For all the compared methods, we use the default settings suggested by the authors. And for the Eq. (2), we take σ^2 = 0.4 which has the best contribution to the results.

The MAE and F-measure evaluation results on both RGBD1* and RGBD2* datasets are shown in Fig. 3 and Fig. 4, respectively. From the comparison results, it can be observed that our saliency detection method is superior and can obtain more precise salient regions than that of other approaches. Besides, the proposed algorithm is the most robust.

The PR curve and ROC curve evaluation results are shown in Fig. 5 and Fig. 6, respectively. From the precision-recall curves and ROC curves, we can see that our saliency detection results can achieve better results on both RGBD1* and RGBD2* datasets.

The visual comparisons are given in Fig. 7, which clearly demonstrate the advantages of our method. We can see that our method can detect both single salient object and multiple salient objects more precisely. Besides, by intermediate results, it shows that by exploiting depth cue information of more layers, our proposed method can get more accurate and robust performance. In contrast, the compared methods may fail in some situations.

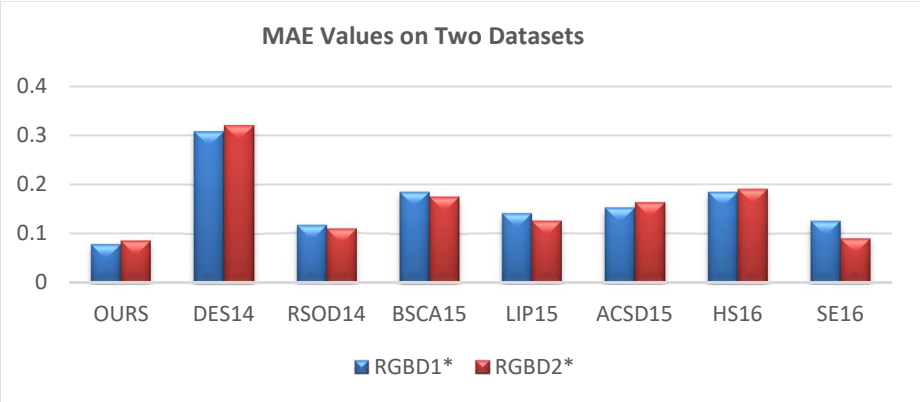

Fig. 3. The MAE results on two datasets. The lower value, the better performance.

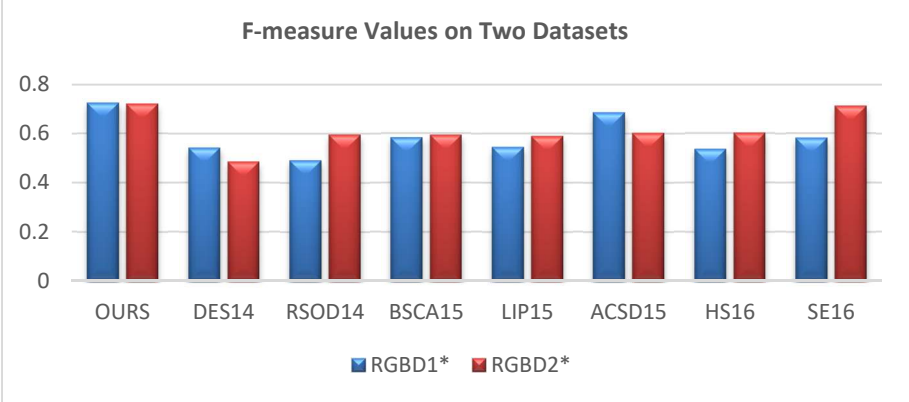

Fig. 4. The F-measure results on two datasets. The higher value, the better performance.

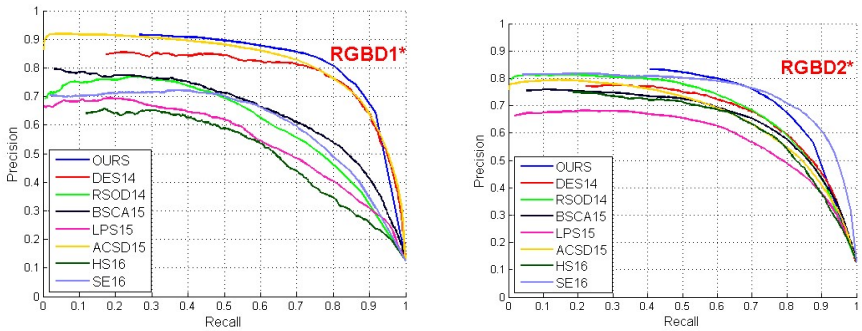

Fig. 5. From left to right: PR curve on RGBD1* dataset and PR curve on RGBD2* dataset.

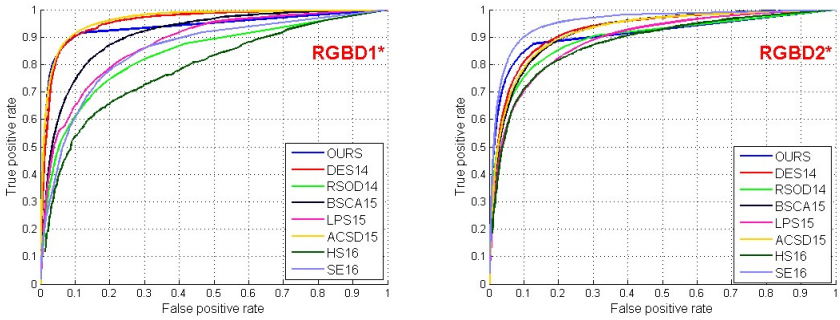

Fig. 6. From left to right: ROC curve on RGBD1* dataset and ROC curve on RGBD2* dataset.

Fig. 7. Visual comparison of saliency map on two datasets, GT represents ground truth.

4 Conclusion

In this paper, we proposed a multilayer backpropagation saliency detection algorithm based on depth mining. The proposed algorithm exploits depth cue information of four layers: in the preprocessing layer, we obtain center-bias map and depth map; in the first layer, we mix depth cue to prominent salient object; in the second layer, we extend depth map to prominent salient object' edges; in the third layer, we reprocess depth cue to eliminate background noises. And the experiments' results show that the proposed method outperforms the existing algorithms in both accuracy and robustness in different scenarios. To encourage future work, we make the source codes, experiment data and other related materials public. All these can be found on our project website: https://chunbiaozhu.github.io/CAIP2017/.

Acknowledgments. 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), the grant of Guangdong Province Projects of 2014B010117007 and the grant of Shenzhen Peacock Plan (No.20130408-183003656).

References

1. 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.
2. 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.
3. Donoser M, Urschler M, Hirzer M, et al. Saliency driven total variation segmentation[C]// IEEE, International Conference on Computer Vision. IEEE, 2009:817-824.
4. Sun J, Ling H. Scale and object aware image retargeting for thumbnail browsing[C]// IEEE International Conference on Computer Vision. 2011:1511-1518.
5. 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.
6. Qin Y, Lu H, Xu Y, et al. Saliency detection via Cellular Automata[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2015:110-119.
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. Murray N, Vanrell M, Otazu X, et al. Saliency estimation using a nonparametric low-level vision model[J]. 2011, 42(7):433-440.
9. 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.
10. 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.
11. Zhu C, Li G, Wang W, et al. Salient Object Detection with Complex Scene based on Cognitive [C]// IEEE Third International Conference on Multimedia Big Data. IEEE, 2017:33-37.
12. Cheng Y, Fu H, Wei X, et al. Depth Enhanced Saliency Detection Method[J]. European Journal of Histochemistry Ejh, 2014, 55(1):301-308.
13. 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.
14. Ju R, Liu Y, Ren T, et al. Depth-aware salient object detection using anisotropic center-surround difference[J]. Signal Processing Image Communication, 2015, 38(C):115-126.
15. Guo J, Ren T, Bei J. Salient object detection for RGB-D image via saliency evolution[C]// IEEE International Conference on Multimedia and Expo. IEEE, 2016:1-6