Salient Object Detection in RGB-D Image Based on Saliency Fusion and Propagation

Jingfan Guo[1,2], Tongwei Ren[1,2,*], Jia Bei[1,2], Yujin Zhu[2]
[1] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
[2] Software Institute, Nanjing University
gjf11@software.nju.edu.cn, rentw@nju.edu.cn, {beijia, zyj12}@software.nju.edu.cn

ABSTRACT

Automatic detection of salient objects in images attracts much research attention for its usage in numerous multimedia applications. In this paper, we propose a *saliency fusion and propagation strategy* based salient object detection method for RGB-D images, in which multiple cues are fused to provide high precision detection result and saliency propagation is utilized to improve the completeness of salient objects. To each RGB-D image, we firstly generate the saliency maps based on color cue, location cue and depth cue independently. Then, we fuse the saliency maps and obtain a high precision saliency map. Finally, we propagate saliency to obtain more complete salient objects. We evaluate the proposed method on two public data sets for salient object detection, NJU400 and RGBD Benchmark. The experimental results demonstrate saliency fusion and propagation are effective in salient object detection and our method outperforms the state-of-the-art methods.

Categories and Subject Descriptors

I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding; I.4.9 [**Image Processing and Computer Vision**]: Applications

General Terms

Algorithms, Human Factors

Keywords

Salient object detection, RGB-D image, multiple cues fusion, saliency propagation

1. INTRODUCTION

Salient object detection aims to detect the attractive objects to human viewers in an image, without any prior knowledge of image content [4]. It is widely used as a fundamental of numerous multimedia applications, including image compression [8], information retrieval [11, 20] and photo editing [3, 18].

In the past years, amounts of salient object detection methods have been proposed. These methods adopt different cues in detection. Color cue is explored in various means, such as color contrast and edge, for human vision system is highly sensitive to color information [4]. Location cue, especially center-bias, is also frequently used to improve saliency detection performance, for people prefer to locate the salient object(s) near the center position when taking a photo [17]. Recently, with the emergence of RGB-D image, depth cue is also used as an effective supplement in salient object detection, for depth perception has an obvious impact in visual attention [9]. However, each cue only provides partial information of salient objects, which may lead to inaccurate detection results when only using a single cue. As shown in Figure 1 (c)-(e), we can find that the saliency maps generated by a single cue may omit some salient regions (Figure 1(c)) or bring in insignificant regions (Figure 1(e)). Hence, it is reasonable to combine multiple cues to improve salient object detection result.

There have been several methods using different cues in salient object detection, for example, combining center-bias with color cues [14] or using color and depth cues together [6]. Yet these methods ignore an important fact that the combination of multiple cues may increase the precision of detection results but decrease the recall, which will obstruct the generation of complete salient objects.

In this paper, we propose a novel salient object detection method for RGB-D images by saliency fusion and propagation. Differing from the existing methods which simply combine several cues together, we obtain a high precision saliency map by fusing the saliency maps generated from multiple cues, and further propagate saliency to enlarge salient regions and improve the completeness of salient objects. The proposed method can raise the performance of the existing methods and efficiently generate high quality saliency maps for RGB-D images. We also evaluate our method on two public data sets, and compare it with the state-of-the-art methods as well as with manually labelled ground truths. The experimental results show that the proposed method obtains better performance than a single cue based methods and other existing salient object detection methods for RGB-D images.

The rest of this paper is organized as follows. In Section 2, we briefly introduce several existing saliency detection methods. Then, we elaborate on the details of our method in Section 3. In Section 4, we present the experiment results.

ICIMCS '15 Zhangjiajie, China

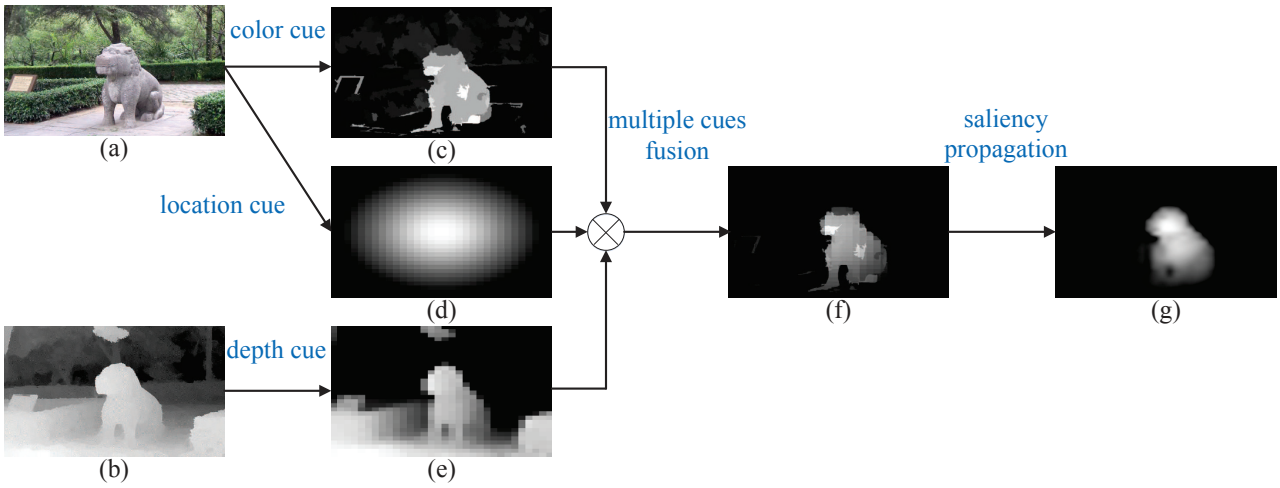

Figure 1: An overview of the proposed method. (a) and (b) RGB channels and depth channel of a RGB-D image. (c)-(e) Saliency maps generated by color cue, location cue and depth cue, respectively. (f) Saliency map generated by multiple cues fusion. (g) Final result after saliency propagation.

Finally, we conclude our work in Section 5.

2. RELATED WORK

We briefly review the salient object detection methods using different cues, including color cue, location cue, depth cue and a combination of multiple cues.

Color cue based methods. Color cue is widely used in salient object detection for its function to human vision system. Achanta *et al.* [1] calculate Euclidean distance between the color of each pixel and the average color of Gaussian blurred image on $L^*a^*b^*$ space. Liu *et al.* [13] formulate the saliency of an object as a global spatial distribution of colors by an assumption that a widely distributed color in an image has low probability to compose a salient object. Cheng *et al.* [4] decompose the image into superpixels and calculate the saliency value of each superpixel based on its contrast to the surrounding superpixels.

Location cue based methods. Location cue is usually used as an effective supplement for salient object detection in natural images. Though there are some arguments about the usage of location cue in salient object detection [12], location cue has been evaluated as effective [17] and used in many outstanding methods. For example, Cheng *et al.* [4] use center-bias to emphasize the saliency values of the superpixels near to image center. Liu *et al.* [14] use border-bias to inhibit the saliency values of the superpixels on image borders.

Depth cue based methods. Depth cue influences salient object detection for the nearer objects attract more human attention [10]. Desingh *et al.* [6] directly use depth map as the input of region contrast algorithm [5] to detect salient objects. Ju *et al.* [9] utilize anisotropic center-surround difference to measure the saliency of each superpixel by how much it outstands from its surroundings on depth map. Peng et al. [16] propose a multi-contextual contrast model including local contrast, global contrast and background contrast to detect salient object from depth map.

Multiple cues based methods. Combining several cues in salient object detection may improve the performance by exploring more characteristics of salient objects from different cues. Niu et al. [15] extend region contrast algorithm for disparity contrast analysis and simultaneously utilize a prior knowledge that salient object is often located in the area of small or zero disparity. Desingh *et al.* [6] adopt region contrast algorithm on RGB image and depth map separately and then fuse these two saliency maps by SVM regression to obtain a final result. Peng *et al.* [16] propose a three-stage approach which consists of low-level feature contrast, mid-level region grouping and high-level prior enhancement. Fang *et al.* [7] combine the feature maps of color, intensity, texture and depth which are extracted from a patch based RGB-D image with an adaptive weighting method. Chen *et al.* [2] propose a patch based method to extend existing 2D saliency detection method with the assumption that a majority of pixels in a patch should have the same depth if the patch belongs to a salient object. Tang *et al.* [19] utilize region contrast algorithm together with depth map to get a noise-filtered salient patch and perform object boundary inference to obtain a refined salient object.

3. MULTIPLE CUES FUSION BASED SALIENT OBJECT DETECTION

Figure 1 shows an overview of the proposed method. We first generate the saliency maps with color, location and depth cues independently. Then, we fuse these saliency maps to obtain a high precision saliency map. Finally, we propagate saliency to enlarge the detected salient regions to improve the completeness of salient objects.

3.1 Saliency Maps Generation Based on Multiple Cues Fusion

Color cue. To generate the saliency map based on color cue, we utilize region contrast method [4] for its high efficiency and effectiveness. It calculates the contrast of each superpixel to its surrounding superpixels with

spatial distance based weights, and provides a full-resolution saliency map. The saliency of each superpixel is calculated as follows:

$$S_C(sp_k) = \sum_{sp_k \neq sp_i} e^{\frac{D_S(sp_k, sp_i)}{-\sigma_C^2}} N(sp_i) D_C(sp_k, sp_i), \quad (1)$$

where sp_k is a superpixel and sp_i is a superpixel surrounding sp_k; $D_S(,)$ is the spatial distance between two superpixels and $D_C(,)$ is the color distance of two superpixels on $L^*a^*b^*$ space; $N(sp_i)$ is the pixel number of sp_i, which is used to indicate the weight of sp_i; σ_C is a normalized parameter, and $\sigma_C^2 = 0.4$ in our experiments as in [4].

Location cue. To generate the saliency map based on location cue, we utilize center bias, which has been widely used to improve salient object detection results. We decompose a image into rectangular patches with the number of $M \times N$, and calculate the saliency value of each patch $p_{m,n}$ as follows:

$$S_L(p_{m,n}) = \frac{1}{2\pi\sigma_L^2} e^{-\frac{(m'-1)^2+(n'-1)^2}{2\sigma_L^2}}, \quad (2)$$

where $m' = \frac{2m}{M+1}$ and $n' = \frac{2n}{N+1}$ represent the normalized coordinate of patch $p_{m,n}$; σ_L is a normalized parameter. In our experiments, M and N are set to 32 for a tradeoff between efficiency and effectiveness, and σ_L^2 is assigned the same value 0.4 as σ_C^2.

Depth cue. To generate the saliency map based on depth cue, we utilize a simple but effective bias similar to center bias, in which the nearer image content is considered to attract more attention. We decompose the image in the same mean as generating S_L, and calculate the saliency value of each patch $p_{m,n}$ as follows:

$$S_D(p_{m,n}) = \frac{1}{2\pi\sigma_D^2} e^{-\frac{(1-d_{m,n})^2}{2\sigma_D^2}}, \quad (3)$$

where $d_{m,n}$ is the normalized depth of $p_{m,n}$, here 1 indicates the nearest and 0 indicates the farthest; σ_D is a normalized parameter. For image content has different scales in imaging plane and depth, we assign $\sigma_D^2 = 0.1$ in our experiments.

Multiple cues fusion. For saliency map generated by a single cue may omit some salient regions and/or bring in insignificant regions, we fuse the saliency maps to obtain the common salient regions in all the saliency maps. In our experiment, we utilize elementwisely dot product in fusion:

$$S_M = S_C \cdot S_L \cdot S_D. \quad (4)$$

3.2 Saliency Propagation

The fused saliency map usually has high precision as a joint result of multiple cues, but it is hard to guarantee the completeness of salient objects for missing the regions insignificant on one or several cues. To solve this problem, we improve the saliency propagation strategy in [17] to enlarge the salient regions and obtain more complete salient objects. In saliency propagation, we decompose the image into patches in the same means as generating saliency maps on location cue and depth cue. When propagating saliency from one patch $p_{m,n}$ to another one $p_{i,j}$, the propagation weight between two patches is defined according to their color similarity and spatial distance in 3D space:

$$\omega(p_{m,n}, p_{i,j}) = \omega_C(p_{m,n}, p_{i,j}) \cdot \omega_S(p_{m,n}, p_{i,j}). \quad (5)$$

Here, $\omega_C(p_{m,n}, p_{i,j})$ is a weight determined by the distance between the average colors of patch $p_{m,n}$ and $p_{i,j}$ on $L^*a^*b^*$ space:

$$\omega_C(p_{m,n}, p_{i,j}) = e^{-\frac{\|c_{m,n} - c_{i,j}\|}{\delta_C^2}}, \quad (6)$$

where $c_{m,n}$ and $c_{i,j}$ are the average colors of $p_{m,n}$ and $p_{i,j}$, respectively; δ_C is a normalized parameter, and $\delta_C^2 = 0.2$ in our experiment. And $\omega_S(p_{m,n}, p_{i,j})$ in Eq. (5) is a weight determined by the spatial distance between patch $p_{m,n}$ and $p_{i,j}$ in 3D space:

$$\omega_S(p_{m,n}, p_{i,j}) = e^{-\left(\frac{(m'-i')^2+(n'-j')^2}{\delta_L^2} + \frac{(d_{m,n}-d_{i,j})^2}{\delta_D^2}\right)}, \quad (7)$$

where (m', n') and (i', j') are the normalized coordinates of patch $p_{m,n}$ and $p_{i,j}$ as defined in Eq.(2); $d_{m,n}$ and $d_{i,j}$ are the average depth values of patch $p_{m,n}$ and $p_{i,j}$, respectively; δ_L and δ_D are the parameters to adjust saliency distribution, and $\delta_L^2 = 0.005$ and $\delta_D^2 = 0.02$ in our experiments.

Based on Eq. (5)-(7), we iteratively update the saliency map by propagating saliency among patches. The iteration is terminated when the average change of saliency map is less than a pre-defined threshold, which equals $\frac{1}{MN}$ in our experiment.

4. EXPERIMENTS

4.1 Data Sets and Experiment Settings

To validate the effectiveness of the proposed method, we evaluate the performance of saliency fusion and propagation, and compare the proposed method with the state-of-the-art methods on two public data sets, NJU400 [9] and RGBD Benchmark [16]. NJU400 data set provides 400 stereo images, in which each left image has its corresponding depth map generated by stereo matching. We treat each left image and its depth map as the color channels and depth channel of an RGB-D image to evaluate the performance of our method in handling relative depth. RGBD Benchmark data set provides 1,000 RGB-D images captured by Microsoft Kinect. We use it to evaluate the performance of our method in handling absolute depth. Both these two data sets provide pixel-level manual-labelled ground truths of salient objects.

We employ precision-recall (PR) curve to represent detection performance. When plotting the PR curve for a certain image, we generate a binary segmentation result of the saliency map by fixed thresholds which vary from 0 to 255. As for PR curve of a whole data set, it is averaged from the curves of all the images in this data set.

The propose method is implemented by Matlab. All the experiments are executed on a computer with 3.4GHz CPU and 8GB memory.

4.2 Experimental Results

We first illustrate the effectiveness of our fusion strategy by comparing with elementwise addition. Figure 3 shows that dot product strategy performs better than addition for it generates salient regions with high precision which can avoid bringing in insignificant regions in the further saliency propagation.

We further evaluate the performance of our propagation strategy by comparing with the propagation strategy in [17]. Figure 4 shows that our propagation strategy outperforms the strategy in [17] for it considers the distances of depth in

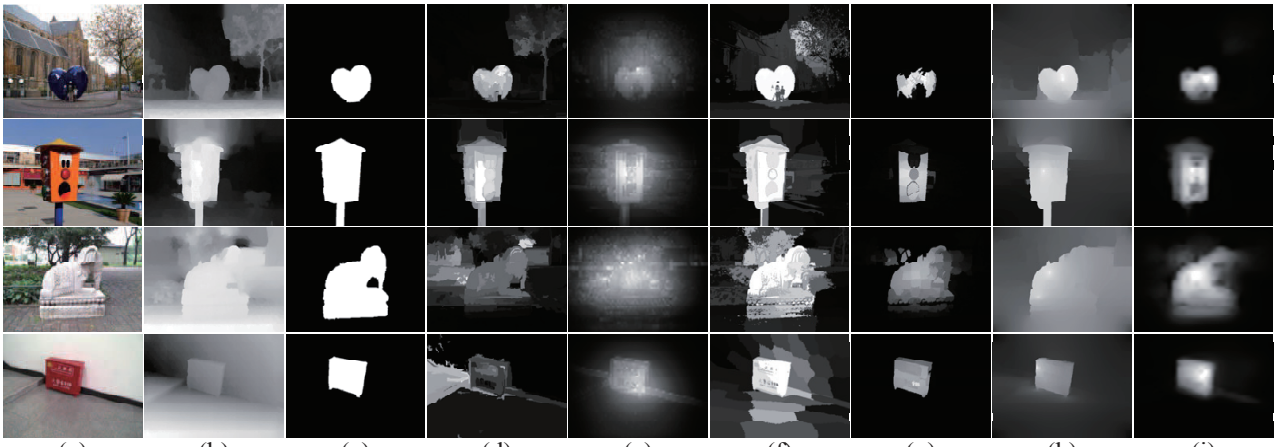

Figure 2: Examples of comparison with the state-of-the-art methods. (a) and (b) RGB channels and depth channels of RGB-D images. (c) Manual-labelled ground truths. (d)-(h) Saliency maps generated by Niu2012[15], Fang2013[7], Cheng2014[5], Peng2014[16] and Tang2015[19], respectively. (i) Our results.

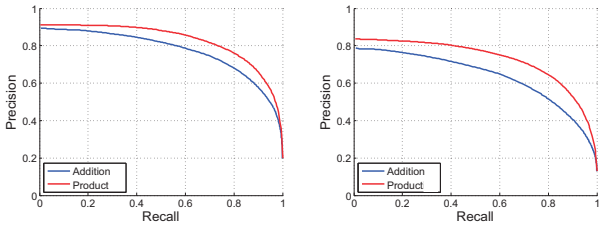

Figure 3: Comparison of fusion strategies. (a) PR curve on NJU400 data set. (b) PR curve on RGBD Benchmark data set.

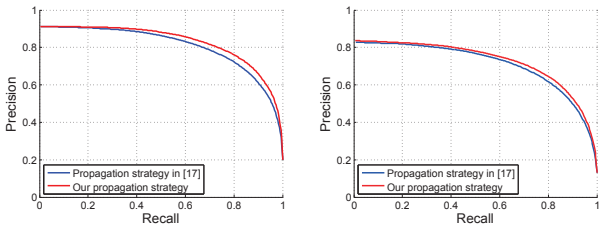

Figure 4: Comparison of propagation strategies. (a) PR curve on NJU400 data set. (b) PR curve on RGBD Benchmark data set.

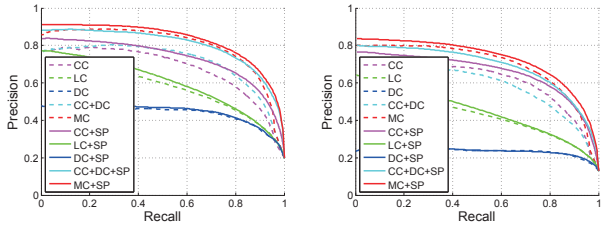

Figure 5: Effectiveness of saliency map fusion and saliency propagation. (a) PR curve on NJU400 data set. (b) PR curve on RGBD Benchmark data set.

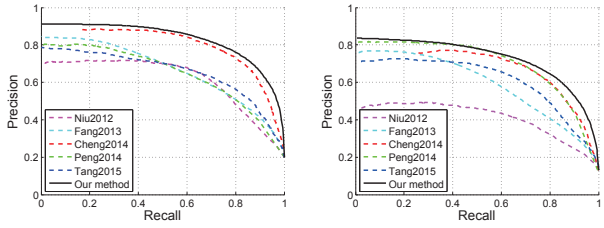

Figure 6: Comparison with the state-of-the-art methods. (a) PR curve on NJU400 data set. (b) PR curve on RGBD Benchmark data set.

saliency propagation to avoid the over-propagation among distant patches in 3D space.

Then we evaluate the effectiveness of saliency fusion and propagation. As shown in Figure 5, saliency maps generated by multiple cues fusion (MC) obviously outperform the saliency detection results based on color cue (CC), location cue (LC), depth cue (DC) and combination of color and depth cues (CC+DC). We can also find that saliency propagation (SP) may improve the salient object detection results and obtain better performance. It means that saliency fusion and propagation are effective and they are indispensable in our method.

We also compare our approach with five existing salient object detection methods for RGB-D images, including Niu2012[15], Fang2013[7], Cheng2014[5], Peng2014[16] and Tang2015[19]. Figure 2 shows some examples of comparison results. We can find that out method obtains good performance on all the example images, but other methods may omit salient regions or bring in insignificant regions on several images. Figure 6 shows the comparison results on NJU400 data set and RGBD Benchmark data set. Our

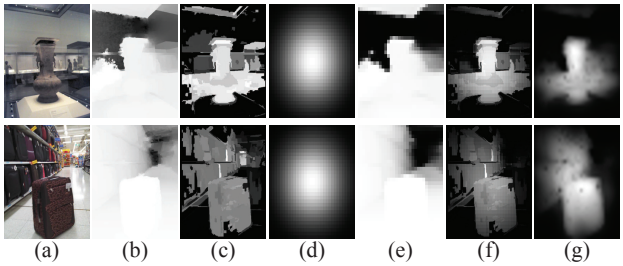

Figure 7: Examples of failure results. (a) and (b) RGB channels and depth channel of RGB-D image. (c)-(g) Saliency maps generated based on color cue, location cue and depth cue, multiple cues fusion, and saliency propagation, respectively.

method obtains better performance than Niu2012, Fang2013 and Tang2015. It has close performance to Cheng2014 on NJU400 data set and Peng2014 on RGBD Benchmark data set, but outperforms them on the other data set, respectively.

4.3 Discussion

In the experiments, we also find some limitations of our method. For example, as shown in the top row of Figure 7, if some insignificant regions are contained in all the saliency maps generated by a single cue, there will be hard to remove them from the final result. And as shown in the bottom row of Figure 7, if some salient region is omitted from one or several saliency maps generated by a single cue, such as the bottom of the suitcase is omitted in saliency maps generated based on color cue and location cue, it may be only partial contained after saliency propagation.

5. CONCLUSION

In this paper, we propose a novel salient object detection method for RGB-D images by saliency fusion and propagation. Differing from generating saliency map using a single cue or simply combining several cues together, our method takes the advantage of the fused saliency map for its high precision and overcomes its shortcoming in incompleteness of salient objects by saliency propagation. The experimental results on two public data sets show that saliency fusion and propagation are effective in salient object detection and our method achieves better performance than existing salient object detection methods for RGB-D images.

6. ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviews for their helpful suggestion. This paper is supported by Natural Science Foundation of China (61202320), Research Project of Excellent State Key Laboratory (61223003), and Natural Science Foundation of Jiangsu Province (BK2012304).

7. REFERENCES

[1] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In *CVPR*, pages 1597–1604. IEEE, 2009.

[2] F. Chen, C. Lang, S. Feng, and Z. Song. Depth information fused salient object detection. In *ICIMCS*, page 66. ACM, 2014.

[3] Y. Chen, Y. Pan, M. Song, and M. Wang. Improved seam carving combining with 3d saliency for image retargeting. *NEUCOM*, 151:645–653, 2015.

[4] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu. Global contrast based salient region detection. *TPAMI*, 37(3):569–582, 2015.

[5] Y. Cheng, H. Fu, X. Wei, J. Xiao, and X. Cao. Depth enhanced saliency detection method. In *ICIMCS*, page 23. ACM, 2014.

[6] K. Desingh, K. M. Krishna, D. Rajan, and C. Jawahar. Depth really matters: Improving visual salient region detection with depth. In *BMVC*, 2013.

[7] Y. Fang, J. Wang, M. Narwaria, P. Le Callet, and W. Lin. Saliency detection for stereoscopic images. In *VCIP*, pages 1–6. IEEE, 2013.

[8] C. Guo and L. Zhang. A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. *TIP*, 19(1):185–198, 2010.

[9] R. Ju, L. Ge, W. Geng, T. Ren, and G. Wu. Depth saliency based on anisotropic center-surround difference. IEEE, 2014.

[10] C. Lang, T. V. Nguyen, H. Katti, K. Yadati, M. Kankanhalli, and S. Yan. Depth matters: Influence of depth cues on visual saliency. In *ECCV*, pages 101–115. Springer, 2012.

[11] L. Li, S. Jiang, Z.-J. Zha, Z. Wu, and Q. Huang. Partial-duplicate image retrieval via saliency-guided visual matching. *IEEE MM*, 20(3):13–23, 2013.

[12] Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille. The secrets of salient object segmentation. In *CVPR*, pages 280–287. IEEE, 2014.

[13] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum. Learning to detect a salient object. *TPAMI*, 33(2):353–367, 2011.

[14] Z. Liu, W. Zou, and O. Le Meur. Saliency tree: A novel saliency detection framework. *TIP*, 23(5):1937–1952, 2013.

[15] Y. Niu, Y. Geng, X. Li, and F. Liu. Leveraging stereopsis for saliency analysis. In *CVPR*, pages 454–461. IEEE, 2012.

[16] H. Peng, B. Li, W. Xiong, W. Hu, and R. Ji. Rgbd salient object detection: A benchmark and algorithms. In *ECCV*, pages 92–109. Springer, 2014.

[17] T. Ren, R. Ju, Y. Liu, and G. Wu. How important is location in saliency detection. In *ICIMCS*, page 10. ACM, 2014.

[18] T. Ren, Y. Liu, and G. Wu. Image retargeting based on global energy optimization. In *ICME*, pages 406–409. IEEE, 2009.

[19] Y. Tang, R. Tong, M. Tang, and Y. Zhang. Depth incorporating with color improves salient object detection. *TVC*, pages 1–11, 2015.

[20] X. Xu, W. Geng, R. Ju, Y. Yang, T. Ren, and G. Wu. Obsir: Object-based stereo image retrieval. In *ICME*, pages 1–6. IEEE, 2014.