Improved Salient Object Detection Based on Background Priors

Tao Xi[1](✉), Yuming Fang[2], Weisi Lin[1], and Yabin Zhang[1]

[1] Nanyang Technological University,
50 Nanyang Avenue, Singapore 639798, Singapore
xitao1989@gmail.com
[2] Jiangxi University of Finance and Economics, Nanchang, China

Abstract. Recently, many saliency detection models use image boundary as an effective prior of image background for saliency extraction. However, these models may fail when the salient object is overlapped with the boundary. In this paper, we propose a novel saliency detection model by computing the contrast between superpixels with background priors and introducing a refinement method to address the problem in existing studies. Firstly, the SLIC (Simple Linear Iterative Clustering) method is used to segment the input image into superpixels. Then, the feature difference is calculated between superpixels based on the color histogram. The initial saliency value of each superpixel is computed as the sum of feature differences between this superpixel and other ones in image boundary. Finally, a saliency map refinement method is used to reassign the saliency value of each image pixel to obtain the final saliency map for images. Compared with other state-of-the-art saliency detection methods, the proposed saliency detection method can provide better saliency prediction results for images by the measure from precision, recall and F-measure on two widely used datasets.

Keywords: Saliency detection · Background priors · Earth movers distance (EMD)

1 Introduction

Digital images have become an increasingly important part of daily life due to the rapid development of information technologies and applications. Various image processing technologies are much desired for visual content analysis in emerging applications. Salient object detection is crucial in various scenarios, such as image retrieval, object recognition, image classification, etc. To effectively locate the visually attractive or interesting objects in images, saliency detection has been widely explored in the research area of computer vision recently [5,6,8,11].

Salient object detection is regarded as a high-level perception process during observers' viewing of visual scenes [4]. The authors of the study [4] labeled rectangle boxes including objects as the ground truth for performance evaluation of salient object detection. They proposed a saliency detection method

Y.-S. Ho et al. (Eds.): PCM 2015, Part I, LNCS 9314, pp. 411–420, 2015.
DOI: 10.1007/978-3-319-24075-6_40

based on machine learning. A new set of features are extracted for salient object prediction by employing conditional random field (CRF). In [5], Achanta et al. exploited the effectiveness of color and luminance features in salient object detection, and proposed an algorithm to detect the salient objects with well-defined boundaries. In [6], Cheng et al. developed a saliency extraction method by measuring regional contrast to create high-quality segmentation masks. In [7], a novel saliency detection model for JPEG images was proposed by Fang et al., in which the DCT difference is used to calculate the saliency degree of each image block. The authors of [8] exploited two common priors of the background in natural images to provide more clues for saliency detection. In [9], Shen and Wu. proposed a novel low rank matrix recovery based saliency detection method by combining high-level priors and the low-level visual features. In [10], Xie et al. built a Bayesian framework based saliency detection method by employing low and mid level visual cues. In [11], Yang et al. proposed a salient object detection method for saliency prediction in images by using graph-based manifold ranking, which incorporates boundary priors and local cues.

Most existing saliency detection models mentioned above calculate the saliency map by computing the feature differences between center-surround patches (or pixels) in images. Among these models, several studies regard the image boundary as the background priors for saliency detection, because it is commonly accepted that photographers would focus on the objects in natural scenes when taking photos [21]. However, boundary information cannot represent the complete background estimation when some of the salient region is located in the image boundary, and this would result in an inaccurate saliency map.

In this paper, an effective saliency detection method is proposed by employing an image segmentation algorithm to adjust the saliency value of image pixels to address the problem mentioned above. First, the input image is divided into superpixels by using the SLIC (Simple Linear Iterative Clustering) method [14], and the color histogram of each superpixel is extracted as the feature. Then the Earth Mover's Distance (EMD) is adopted to calculate the feature differences between superpixels because it was widely used in computer vision as an effective similarity metric. The saliency degree for each superpixel is computed as the sum of feature differences between this superpixel and other superpixels within the image boundary (which is regarded as the background priors), since saliency is the contrast between current region and background. Finally, a graph-based image segmentation method (GS04) [13] is employed to adjust the saliency value of each image pixel for the final saliency map prediction, since the segmented superpixels by SLIC are almost uniform size and are more suitable to used as the basic processing units for image representation, while in the segment regions from GS04, much more object boundaries are reserved [14]. Experimental results have demonstrated that our saliency detection method can obtain better performance for salient object detection than other existing ones.

2 Improved Saliency Detection Based on Background Priors

We provide the framework of the proposed saliency detection method in Fig. 1. Given an image, we first segment it into superpixels as computational units, and calculate the initial saliency map by measuring the dissimilarity between the superpixel and those in the image boundary. Then, the background prior is updated for the saliency map calculation. Finally, we adjust the saliency value of each image pixel by utilizing a graph-based segmentation method (GS04) [13] to obatain the final saliency map of the input image.

2.1 Pre-processing

In this study, the SLIC method [14] is employed to divide the image into superpixels. In many vision tasks, superpixels are used to capture image redundancy and reduce the complexity of image processing algorithms. Also, compact and highly uniform superpixels with respect to image color and contour are much desired. The authors of [14] compared five state-of-the-art superpixel extraction methods with respect to speed, image boundaries, memory efficiency, and the performance on segmentation results. Compared with other existing methods (include GS04), the SLIC algorithm is easy to be implemented and it provides compact and highly uniform superpixels [14]. Here, we use the SLIC algorithm to generate superpixels as computational units for saliency calculation. For each superpixel, we extract the color histogram as its feature. To improve the computational efficiency, a sparse histogram representation is adopted for feature extraction of superpixels here. Specifically, we first use a color quantization method [16] to reduce quantization bins of each channel in RGB color space. After that, we transform RGB color space to CIEL*A*B* color space due to the latter's perceptual property.

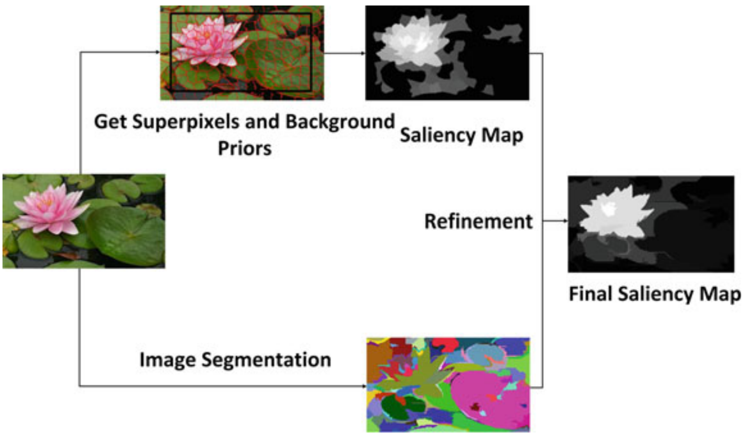


Fig. 1. The framework of the proposed salient object detection method.

2.2 Initial Saliency Map Calculation

Generally, photographers always focus on salient objects when taking photos. Thus, the foreground objects are more likely to be located at the image center. Due to this characteristic, "Center Bias" mechanism can be used as a significant factor in saliency detection. Existing studies have also pointed out that photographers always tends to place interested objects near the center of the shot and to enhance foreground objects relative to the background [12].

In this study, we regard image boundary as the background priors for saliency detection. Although some part of an image boundary might not be background, the image boundary usually captures the representative statistics of background region for an image. For a superpixel S_i where i is the location index of the superpixel, we measure its saliency degree by computing its color difference to all superpixels in the image boundary. Denoting the height and width of the current image as H and W (as shown in Fig. 2), we define the outside of the black rectangle box as $BackgroundPriors$ region (BP region), the inside of the black rectangle box is defined as $Non\text{-}BackgroundPriors$ region ($N\text{-}BP$ region). Specifically, we define the width of four sides of the BP region as $D_{top} = H/n$, $D_{bottom} = H/n$, $D_{left} = W/n$, $D_{right} = W/n$, respectively, n is a parameter which is greater than 1 to control the size of the BP region (here, we set n=30). For each superpixel S_i, if there is any one pixel falling into BP region, this superpixel is labeled as $BackgroundPriors$ superpixel (BP superpixel). Otherwise, this superpixel is labeled as $Non\text{-}BackgroundPriors$ superpixel ($N\text{-}BP$ superpixel). For each superpixel S_i, its saliency value $Sal(S_i)$ is defined as:

$$Sal(S_i) = \left(\sum_{j=1}^{M} EMD\left(S_i, S_j\right) \right)^{\frac{1}{M}}, \quad S_j \in BP\ superpixels \tag{1}$$

where M is the number of BP superpixels in the image; EMD is earth mover's distance (EMD), which is a distance metric between the color histograms of S_i and S_j. In [6], the histogram distance is measured by computing the sum of the

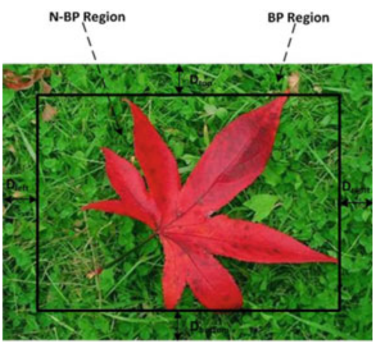


Fig. 2. The illustration of BP region and N-BP region.

differences between each pair bins. In this study, we use the EMD [16] to calculate the differences between superpixels, because it can obtain better performance for perceptual similarity measure than other similarity metrics [16]. Generally, EMD can be regarded as a solution to the famous transportation problem in linear optimization. For two different color histograms, the EMD is the minimum cost of changing one histogram into the other.

2.3 Saliency Map Refinement

Several recent approaches exploit the background priors to enhance saliency computation and their experimental results demonstrate that the background priors is effective in saliency detection. However, they simply treat all image boundaries as the background region. These models might fail if the salient object is slightly overlapped with the boundary due to the fact that they simply use all image boundary as the background priors [18]. To address this drawback, the initial saliency map is refined by reassigning the saliency value of each image pixel. Specifically, we first segment the input image into different regions by employing a graph-based image segmentation method (GS04) [13], and then reassign saliency value to each pixel by computing the mean value of saliency values in each new region. Compared with SLIC, the segment regions from GS04 are more likely to include the integrated objects than the results from SLIC [14]. Thus, the saliency value in foreground or background from GS04 can be more uniform than that of SLIC. The saliency map with refinement can be computed as follows.

$$S_p = \frac{1}{N_p} \sum Sal_{i,j}, \quad (i,j) \in reg_p \tag{2}$$

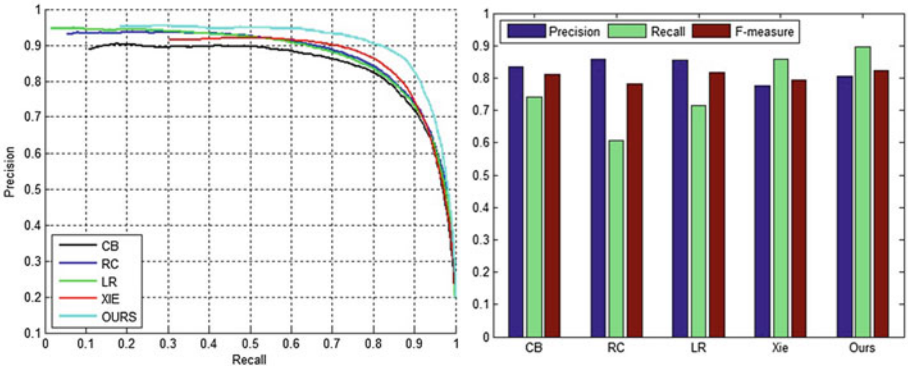
where reg_p denotes the p_{th} region from the segmentation result, N_p represents the number of image pixels in reg_p. We can obtain the final saliency map by normalization as follows.

$$Norm(Sal_{i,j}) = \frac{Sal_{i,j} - Sal_{min}}{Sal_{max} - Sal_{min}} \tag{3}$$

where Sal_{min} and Sal_{max} are the minimum and maximum values of $Sal_{i,j}$ over all pixels in the image.

3 Experiments

In this section, we evaluate the performance of our saliency detection method and analyze the effectiveness of the proposed saliency map refinement method on two public datasets: ASD [5] and MSRA [15]. ASD includes 1000 original images with their corresponding binary masks of salient objects, while there are 5000 images and the corresponding ground truth in MSRA. The images in ASD are relatively simpler than the ones in MSRA. The images from MSRA contain complex background and low contrast objects, which make this dataset challenging for saliency detection. For single image saliency detection, we compare our



(a) (b)

Fig. 3. Comparison results between the proposed saliency detection model and other four state-of-the-art methods on ASD dataset. (a) *P-R* curve. (b) Mean *Precision*, *Recall*, and *F-measure* values of the compared methods.

saliency detection method with four state-of-the-art methods of saliency detection: RC [6], CB [19], XIE [10], and LR [9]. Although there have been dozens of saliency detection methods designed in recent decades since the well-known saliency detection method [20] was proposed by Itti et al., a lot of methods aim to predict humans? eye fixation and cannot get promising results for salient object detection. Therefore, we choose several recent representative methods for salient object detection models to conduct the comparison experiment. Also, for each saliency detection method, we set the default parameters provided by the authors to run the executable source codes.

The *Precision*, *Recall* and *F-measure* are employed to evaluate the matching degree between a saliency map and its corresponding ground truth. *Precision -Recall* curve (*P-R* curve) is commonly applied to measure the performance of the salient object detection model. The *P-R* curve can be drawn as the curve of *Precision* versus *Recall* by means of setting different thresholds. Generally, a higher *P-R* area means a better predication performance for a specified saliency detection method. Furthermore, similar with [5], we use the image dependent adaptive thresholds to evaluate the performance of our model. The image dependent adaptive threshold [5] is defined by as twice of the mean of the saliency map. The *F-measure* is introduced to evaluate the performance of saliency detection models when using image dependent adaptive threshold. The *F-measure* is defined as:

$$F = \frac{(1+\beta^2)\cdot Precision \cdot Recall}{\beta^2 \cdot Precision + Recall} \tag{4}$$

where the coefficient β^2 is set to 1, indicating the equal importance of *Precision* and *Recall* as [17].

3.1 Performance Evalation on ASD and MSRA Datasets

Images in the ASD dataset contains objects with high contrast and clear contour. Figure 3 (a) shows the P-R curve of RC [6], CB [19], XIE [10], LR [9] and our method, we can see that our method has the highest P-R curve and thus outperforms all the other methods. Also, we compare the performance between our method and all the other methods with the adaptive thresholds for each image. As shown in Fig. 3 (b), our method also obtains the best performance among the compared models on this dataset. We provide some visual comparison samples from the ASD dataset in Fig. 4. From this figure, it is apparent that the saliency map generated by our method on these samples are more consistent with the ground truth compared with other existing models. Specifically, our method outperforms other methods when the visual scene is complex (See the first two rows of Fig. 4).

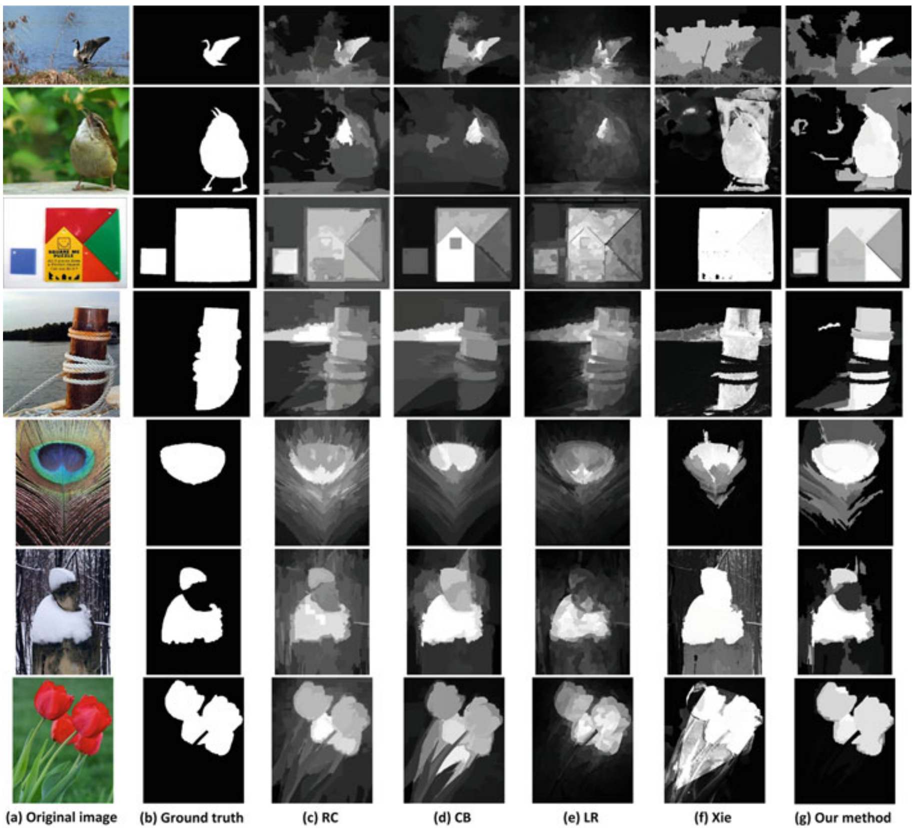
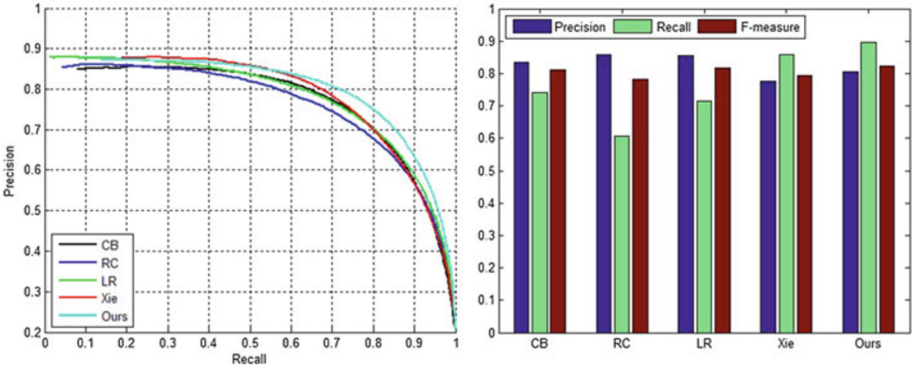


Fig. 4. Visual comparison of our saliency detection method with four other methods.



(a) (b)

Fig. 5. Comparison results between our saliency detection method and four state-of-the-art methods on MSRA dataset. (a) *P-R* curve. (b) Mean *Precision*, *Recall*, and *F-measure* values of the compared methods.

In the MSRA dataset, there are some images including objects with low contrast, which make this dataset more challenging. The experimental results from the compared models on this database are shown in Fig. 5. From Fig. 5(a), our method yields higher *P-R* curve than CB, RC, LR and XIE for the MSRA dataset. Furthermore, we can see that our method has a competitive *F-measure* in Fig. 5(b), when saliency maps are generated by the adaptive thresholds. In general, our method is able to better detect salient objects in images.

3.2 Effectiveness of Saliency Map Refinement

The saliency map refinement is an important contribution in the proposed method. We test the performance of our method with refinement and without refinement. As shown in Fig. 6, the performance of our method without the refinement decreases obviously. To intuitively demonstrate the effectiveness of refinement mechanism, Fig. 7 gives some saliency samples of our method with refinement and without refinement. For the images shown in Fig. 7(a), the salient

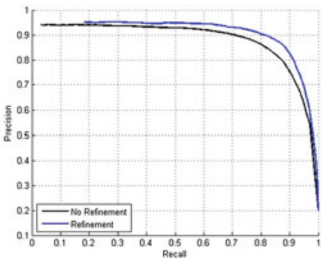


Fig. 6. Evaluation on the effectiveness of refinement.

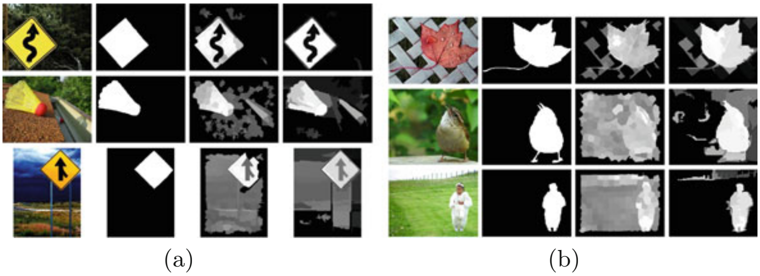


Fig. 7. Visual comparison of our saliency detection method with and without refinement on integrity (a) and uniformity (b) of saliency values on salient object. From left to right (in (a) and (b)): input image, ground truth, saliency map generated by our method without refinement, saliency map generated by our method with refinement.

objects are near to the image boundary, which results in incompleteness of detected object (the third column). By employing refinement, we can obtain an complete salient object in the saliency map (the fourth column), since the segmentation results of GS04 are more likely to include the integrated objects. On the other hand, the refinement can make saliency values of image pixels in the salient objects more uniform, and it tends to highlight the salient object from the image due to the use of GS04, as shown in Fig. 7(b). This strongly indicates that refinement is a significant contributor to the effectiveness of our method.

4 Conclusion

We have presented a new salient object detection method based on background priors, in which the saliency is measured by EMD between current superpixel and boundary superpixels. To address the drawback when regarding the whole image boundary as background priors and salient objects are overlapped with the boundary, we design a saliency map refinement method to reassign saliency value by exploiting the segmentation results of a graph-based image segmentation algorithm. By employing this method, we find that the proposed method can highlight the salient objects more integrally and uniformly. Experimental results on two public test datasets have demonstrated that the proposed saliency detection method outperforms the state-of-the-art saliency detection approaches for salient objects prediction.

Acknowledgements. This research was supported by Singapore MOE Tier 1 funding (RG 36/11: M4010981), and the Rapid-Rich Object Search (ROSE) Lab at the Nanyang Technological University, Singapore. The ROSE Lab is supported by the National Research Foundation, Prime Ministers Office, Singapore, under its IDM Futures Funding Initiative and administered by the Interactive and Digital Media Programme Office.

References

1. Chen, L., Xie, X., Ma, W., Zhang, H., Zhou, H.: Image adaptation based on attention model for small-form-factor devices. In: ICME (2003)
2. Ouerhani, N., Bracamonte, J., Hugli, H., Ansorge, M., Pellandini, F.: Adaptive color image compression based on visual attention. In: ICIAP (2001)
3. Stentiford, F.: A visual attention estimator applied to image subject enhancement and colour and grey Level Compression. In: ICPR (2004)
4. Liu, T., Sun, J., Zheng, N., Tang, X., Shum, H.: Learning to detect a salient object. In: CVPR (2007)
5. Achanta, R., Hemami, S.S., Estrada, F.J., Ssstrunk, S.: Frequency tuned salient region detection. In: CVPR (2009)
6. Cheng, M., Zhang, G., Mitra, N., Huang, X., Hu, S.: Global contrast based salient region detection. In: CVPR (2011)
7. Fang, Y., Chen, Z., Lin, W., Lin, C.: Saliency detection in the compressed domain for adaptive image retargeting. IEEE Trans. Image Process. **21**(9), 3888–3901 (2012)
8. Wei, Y., Wen, F., Zhu, W., Sun, J.: Geodesic saliency using background priors. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part III. LNCS, vol. 7574, pp. 29–42. Springer, Heidelberg (2012)
9. Shen, X., Wu, Y.: A unified approach to salient object detection via low rank matrix recovery. In: CVPR (2012)
10. Xie, Y., Lu, H., Yang, M.: Bayesian saliency via low and mid level cues. IEEE Trans. Image Process. **22**(5), 1689–1698 (2013)
11. Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M.: Saliency detection via graph-based manifold ranking. In: CVPR (2013)
12. Tseng, P., Carmi, R., Cameron, I., Munoz, D., Itti, L.: Quantifying center bias of observers in free viewing of dynamic natural scenes. J. Vis. **9**(7:4), 1–16 (2009)
13. Felzenszwalb, P., Huttenlocher, D.: Efficient graph-based image segmentation. Int. J. Comput. Vis. **59**(2), 167–181 (2004)
14. Achanta, R., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: Slic superpixels. Technical report, EPFL. Technical report: 149300(3) (2010)
15. Jiang, H., Wang, J., Yuan, Z., Wu, Y., Zheng, N., Li, S.: Salient object detection: a discriminative regional feature integration approach. In: CVPR (2013)
16. Grauman, K., Darrell, T.: Fast contour matching using approximate earth movers distance. In: CVPR (2004)
17. Liu, Z., Zou, W., Meur, O.: Saliency tree: a novel saliency detection framework. IEEE Trans. Image Process. **23**(5), 1937–1952 (2014)
18. Zhu, W., Liang, S., Wei, Y., Sun, J.: Saliency optimization from robust background detection. In: CVPR (2014)
19. Jiang, H., Wang, J., Yuan, Z., Liu, T., Zheng, N., Li, S.: Automatic salient object segmentation based on contex and shape prior. In: BMVC (2011)
20. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. **20**(11), 1254–1259 (1998)
21. Tian, H., Fang, Y., Zhao, Y., Lin, W., Ni, R., Zhu, Z.: Salient region detection by fusing bottom-up and top-down features extracted from a single image. IEEE Trans. Image Process. **23**(10), 4389–4398 (2014)