ORIGINAL ARTICLE

Saliency detection with color contrast based on boundary information and neighbors

Min Xu · Hanling Zhang

Abstract Object-level saliency detection is significant in many computer vision tasks. In this paper, we propose a novel saliency detection model based on color contrast and image boundaries. The saliency of an image is defined as the contrast between the image elements (regions) and image boundaries elements (regions). We consider the saliency in two-stage procedure rather than in one stage. First of all, according to the definition of saliency, we take four boundaries of image into consideration respectively to obtain a combination coarse saliency map. Furthermore, a new energy function based on the coarse saliency map is proposed, which takes the coarse saliency map as input to yield the final full resolution saliency map. Experimental results on two public datasets demonstrate that the proposed model performs better than the state-of-the-art methods.

Keywords Visual saliency · Color contrast · Energy function · Saliency map

1 Introduction

The human visual system can locate the most important regions in a scene rapidly and accurately [1,2]. Such image regions, which could attract more attention than other regions in the image, are said to be salient regions. The visual saliency detection to model biological visual systems has acquired extensive attention from computer vision researchers and

M. Xu (✉) · H. Zhang
College of Information Science and Engineering,
Hunan University, Changsha 410000, China
e-mail: xumin19891030@163.com

H. Zhang
e-mail: jt_hlzhang@hnu.edu.cn

psychologists [3,4], and many saliency computational models have been proposed to reduce the information redundancy in recent years. As the preprocessing operations, saliency detection were applied to numerous computer vision tasks, such as interesting regions detection [5], image categorization [6], object localization [7] and image compression [8,9]. Although existing saliency models have achieved promising results, some particular limitations remain to be overcome. Figure 1 shows some problems in saliency detection.

The saliency measure, which is the dissimilarity of an object to its surroundings, describes the ability of the object stands out from the image. The salient object will highlight from the scene and will get the prior attention from our visual system. From the perspective of the information processing, saliency algorithms in computer vision can be categorized into bottom-up (data driven) measure [3,10–23] and top-down (task driven) measure [24,25]. The top-down saliency detection algorithms are often associated with a particular task. For instance, in target detection, the calculated saliency map denotes the possible location of the target in the image. Moreover, the bottom-up algorithms, which are more significant, are always based on the low level visual information and do not associate with a specific target.

In this paper, we focus especially on the bottom-up saliency algorithm. The saliency detection algorithm usually generates a saliency map where each value indicates the saliency likelihood of each pixel in the image. In the frequency domain, Hou et al. [17] proposed a simple and fast algorithm which called spectrum residual (SR) using the Fourier Transform. The paper argues that the SR corresponds to image saliency. Referring to the SR, the phase spectrum of the Fourier transform (PFT) and PQFT [8] are presented. The PFT can achieve nearly the same property with SR, while the PQFT combines more features to generate saliency map using the quaternion Fourier Transform.
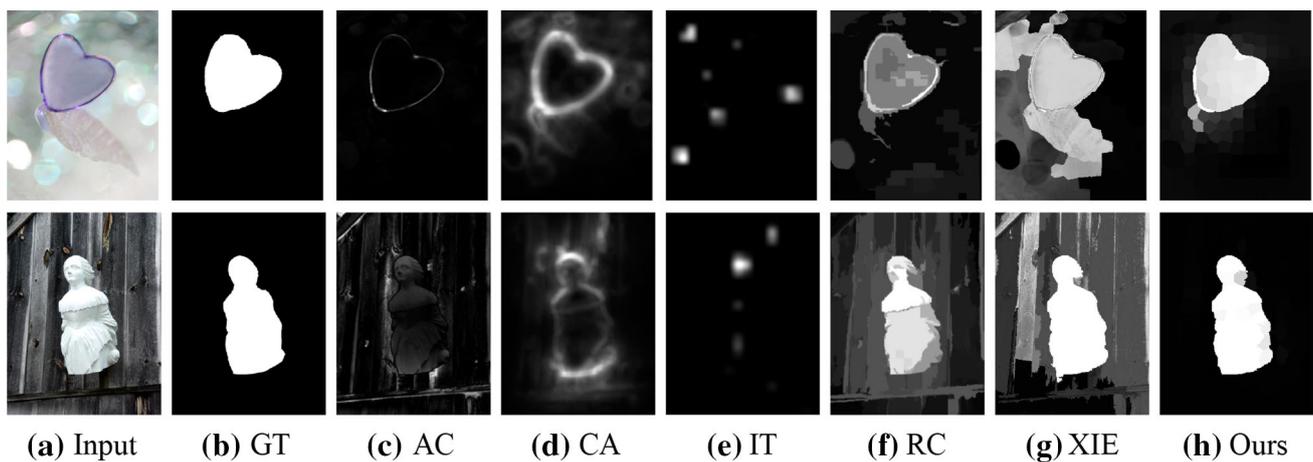
Fig. 1 The saliency maps of different methods. We can see that some methods like AC [10] and CA [14] can not emphasize the whole region or object uniformly, which more highlight the object boundary. And some methods like IT [3] and RC [13] respond more to unrelated stimuli

Li et al. [26] introduced an approach, which based on the scale-space analysis in the frequency domain, models the non-saliency regions with Hyper complex Fourier Transform. On the other hand, many algorithms perform encouragingly in spatial domain. Using the color, intensity and orientation pyramid to model the image saliency, Itti and Koch's [3] proposed a saliency model which proved to be a milestone in saliency detection. Achanta et al. [11] developed a frequency-turned method that defines the saliency of a pixel as its color difference from the average value of all pixels in image. Although this method is simple and efficient, it can not detect the salient object accurately in a complex scene. Using local computation, Hare et al. [16] presented a graph-based solution to obtain a saliency map. Liu et al. [24] did saliency detection using multi-scale measure via linear combination with Gaussian image pyramid. Based on the global contrast, Cheng et al. [13] designed a saliency detection approach which involves either the color contrast or spatial coherence. By exploiting low and mid level cues, Xie et al. [27] suggested a bottom-up model that is in view of the Bayesian framework. Wei et al. [28] defined the saliency of an image region as its shortest path to the virtual node on the image boundary, considered the image boundary as the possible background. In this work, we generate a full resolution saliency map for each input image.

Inspired by [28], we use, for each image in this work, super-pixels which were generated by the SLIC algorithm [29] to measure region saliency. We also define the saliency of a super-pixel as its color relevance to image boundary, and propose a two-stage scheme for our model. In the first stage, we directly define region saliency as the color contrast of a super-pixel to the average of each side of the image, and the super-pixel is salient when there is a high contrast. Then the four calculated maps are then integrated to generate a coarse saliency map. In the second stage, motivated by the matting

method [30] and the coarse saliency map, we introduce a new energy function. Given the color relevance, it takes the coarse saliency map as input, and generates the final saliency map as output. Some visual saliency effects of the proposed method are shown in Fig. 2.

The contribution of this paper is threefold: (1) A new saliency model has been proposed, (2) a coarse saliency map has been generated through the effectively using color contrast on boundaries, (3) a new energy function which takes the coarse saliency map as input has been presented.

The remaining of this paper is organized as follows. Section 2 describes how to generate the coarse saliency map according to the color difference between a super-pixel and the average of super-pixels locating on the four boundaries of image. In Sect. 3, we present a new energy function and in Sect. 4, we analyze the experimental results. Conclusion and the future work are discussed in Sect. 5.
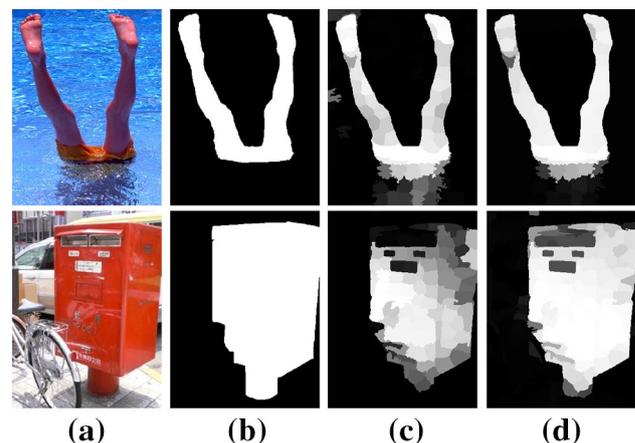

Fig. 2 The visual saliency effects of the proposed model. **a** Input images. **b** Ground truth. **c** Results of the first stage. **d** Results of the second stage
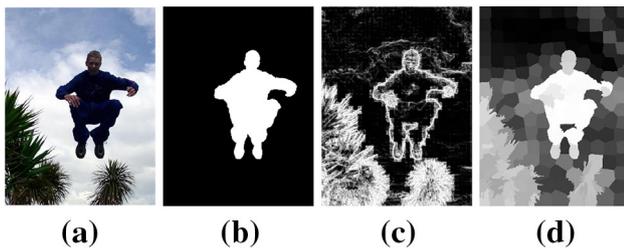
(a) **(b)** **(c)** **(d)**

Fig. 3 **a** Input image. **b** Ground truth. **c** The saliency map by using the contrast between pixel and its neighboring pixels. **d** The saliency map by using the contrast between pixel and image boundary. We can see that our method has a better effect

2 The coarse saliency map

Psychophysical studies show that human attention favors central regions of natural image [31]. The concept of center bias, meaning that image center is more likely to contain salient objects than other regions, is used in many previous works. Those works highlight the object which locates in the image center effectively. However, the effectiveness of the algorithms is limited. The salient object almost seldom touches the image boundary, while the background regions can be easily connected to image boundaries [28]. Therefore, we define the saliency of a super-pixel as its color contrast to the boundaries super-pixels. A higher contrast refers to a higher salient value. We use the boundaries super-pixels instead of neighboring pixels to calculate the contrast. As shown in Fig. 3, because the object boundaries usually have the higher contrast, the approach using the contrast between current pixel and its neighboring (eight areas) pixels not only highlights the boundary of the object but also responds to numerous unrelated visual stimuli. Our result proved to be better when we use the boundaries super-pixels instead of the neighboring pixels to calculate the contrast.

Based on the definition of saliency, we first use the segmentation method that is discussed in [29] to segment the image into N super-pixels. Then the saliency value $v(i)$ of a super-pixel can be defined as:

$$v(i) = 1 - \frac{1}{k} \sum_{j=1}^{k} w_{ij}, \tag{1}$$

where k is the total number of super-pixels on the four boundaries of image. w_{ij} is the color contrast of two super-pixels which is defined as:

$$w_{ij} = e^{-\frac{\|c_i - c_j\|}{\sigma^2}} \quad (i, j \in 1, 2, \ldots, N), \tag{2}$$

where c_i and c_j denote the mean color value of all pixels in a super-pixel in the Lab color space. Actually, the color distance of two super-pixels has appeared in many saliency detection literatures. Moreover, σ represents a constant that controls the strength of the distance.
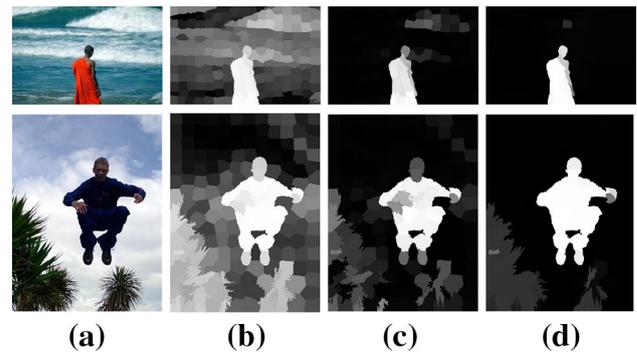

(a) **(b)** **(c)** **(d)**

Fig. 4 Saliency maps in different handing methods. **a** Input images. **b** Saliency maps of consider the four image boundaries integrally. **c** Saliency maps about considering the four boundaries of each image respectively. **d** The final saliency map of the proposed method

The illustration of the saliency maps referencing the four boundaries of image integrally are shown in Fig. 4b. We note that if there is an object contacting to the boundaries (the first row of Fig. 4a) or there exists a boundary which has high contrast with the three others (the second row of Fig. 4a), the final results may not only highlight the object, but also present the irrelevant background (Fig. 4b). That is because the regions between different backgrounds cannot be easily connected, though the internal of backgrounds has highly connectivity. Therefore, though for humans that the tree and sky regions in the second row of Fig. 4a are homogeneous by themselves, the saliency value may be quite different according to the definition of the saliency.

In order to suppress the background effectively, we propose a new solution that considers the four boundaries respectively. Let take the top boundary of the image as an example, for a super-pixel i, by measuring its color contrast to the top boundary, we define its saliency value $v_t(i)$ as:

$$v_t(i) = 1 - \frac{1}{k} \sum_{j=1}^{k} w_{ij} \quad (i \in 1, 2, \ldots, N), \tag{3}$$

where k is the number of super-pixels on the top boundary. Similarly, the saliency value of down boundary, left boundary and right boundary can also be computed with the same method. Then we can obtain four maps: v_t, v_d, v_l and v_r respectively. Finally, we get a coarse saliency map by integrating the four maps according to the following equation:

$$v = v_t \times v_d \times v_l \times v_r. \tag{4}$$

From the visual comparison as shown in Fig. 4b, c, we note that we can suppress the backgrounds to some extent by considering the four boundaries of image respectively. There are two reasons which allow us to consider the four boundaries of image respectively and can contribute to a better result. Firstly, the super-pixels on different image boundaries should have large distance and the appearances are often dissimilar. Secondly, it reduces the inaccurate super-pixels,

i.e., the object connects with the image boundary. Although some salient regions of the object probably do not have excellent appearance, they could be identified when the coarse saliency map is obtained by integrating those four maps. The coarse map gives sufficient cues to the process of the second stage about saliency detection.

3 A new energy function

Due to the lack of sufficient priors or high level knowledge, most saliency detection models cannot achieve good performance. Most bottom-up saliency methods rely on the fundamental assumption that a salient pixel or region presents great dissimilarity with surroundings. The color contrast guides a way to reach this, but it is insufficient. In this section, we design a sparsely connected graph $G = (V, E)$, where V is the super-pixel set in the image and E is the undirected link set. In this work, we realize that each super-pixel is not only relevant to its neighboring super-pixels, but also related to the super-pixels that share common boundaries with its neighboring super-pixels, i.e. the neighbor's neighbor. This encourages the neighboring super-pixels in the image to take the similar value, like the use of local smoothness constraint in many segmentation works.

According to the Sect. 2, we can not only obtain the coarse saliency map, but also the saliency value of each super-pixel. Inspired by the matting method [30], the user is asked to provide a trimap that has given certain foreground and background. There would be an outstanding result according to original image and trimap. Based on the coarse saliency map, we note that if the saliency value of a pixel is high, the possibility that the pixel belongs to the salient pixel is also high, and vice versa. Therefore, based on the coarse saliency map and the local smoothness properties, a novel energy function which can format global optimization is proposed:

$$s^* = \arg\min_s \frac{1}{2} \left(\sum_{i,j=1}^{N} w_{ij}(s_i - s_j)^2 + \mu \sum_{i=1}^{N} \phi_i(s_i - v_i)^2 + \lambda \sum_{i=1}^{N} T_i(s_i - z_i)^2 \right), \quad (5)$$

There are three items on the right side of Eq. (5). Where s can be viewed as a vector $s = [s_1, s_2, \ldots, s_N]^T$, s_i and s_j indicate the saliency for each super-pixel. N represents the number of super-pixels whereas w_{ij} is the weight of two linked super-pixels i and j, which represents the same color contrast as in Eq. (2). The first item indicates that a good saliency map should not change too much between nearby super-pixels.

In the second item, the normalized v_i denotes the saliency value of each super-pixel in the coarse saliency map. In this item, ϕ_i is an important parameter which can be expressed as:

$$\phi_i = e^{-v_i(1-v_i)/\sigma_1^2}. \quad (6)$$

Specifically, this function denotes that if a super-pixel's value is closer to 1 or 0 in the coarse saliency map, the super-pixel has higher possibility to be the certain foreground or background. i.e., it has a more significant impact on the final saliency. The parameter μ controls the importance of the coarse saliency map. The larger the parameter μ, the higher weight the second item has. As a consequence, the final saliency result is more similar with the coarse saliency map (as shown in Fig. 5).

We have obtained a coarse saliency map in the discussion of Sect. 2, and we consider that it can approximately represent the salient object. Affected by the matting solution [30] which used the trimap, we suppose that we can obtain the certain foreground pixels or background pixels using a threshold based on the coarse saliency map. Thus in the third item of Eq. (5), we assume that T_i denotes the certain pixels:

$$T_i = \begin{cases} 1, & \text{if}(v_i \geq \alpha M) \\ 1, & \text{if}(v_i < \beta M) \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

where M represents the mean saliency value of the coarse saliency map. If $v_i \geq \alpha M$, the super-pixel i belongs to the foreground. If $v_i < \beta M$, it belongs to the background. Parameter λ controls the balance of the Eq. (5). The second item and third term mean that the final result should not differ too much from the coarse saliency map.

The minimum solution is computed by setting the derivative of the above energy function to be zero. The resulting solution of Eq. (5) is derived as:
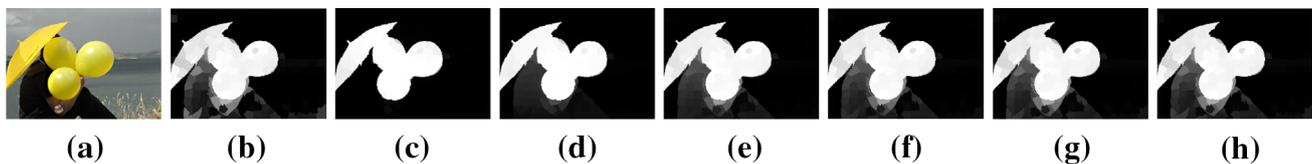

Fig. 5 Illustration saliency maps of different weights μ. **a** Input image. **b** The coarse saliency map. **c–h** Refined saliency maps with different parameters $\mu = 0.1, 10, 50, 100, 150, 200$. With the parameter μ increases, the refined map is similar with the coarse saliency map. The map with $\mu = 0.1$ is the best

$$s^* = (\mathbf{D} - \mathbf{W} + \mu\mathbf{\Psi} + \lambda\mathbf{T})^{-1}(\mu\mathbf{\Psi}V + \lambda\mathbf{T}z), \qquad (8)$$

where $\mathbf{W} = [w_{ij}]_{N \times N}$ is the color relevance matrix, $\mathbf{D} = \mathrm{diag}\{d_{11}, d_{22}, \ldots, d_{NN}\}$ denotes a diagonal matrix and d_{ii} is the sum of the column vector of the color matrix $\mathbf{W}$, which can be expressed as $d_{ii} = \sum_j w_{ij}$. $\mathbf{\Psi} = \mathrm{diag}\{\phi_1, \phi_2, \ldots, \phi_N\}$ is a diagonal matrix that indicates the value of ϕ_i, $\mathbf{T} = \mathrm{diag}\{T_1, T_2, \ldots T_N\}$ denotes the certain foreground and background pixels. $V = [v_1, v_2, \ldots, v_N]^T$ is the saliency value of the coarse saliency map which we have obtained in Sect. 2 whereas $\mathbf{z} = [z_1, z_2, \ldots z_N]^T$ represents the vector of certain foreground pixels.

It is apparent that $\mathbf{D}$ and $\mathbf{W}$ can breezily be obtained according to their definition, and both of them are sparse matrix. Since the number of super-pixels is small, the calculated load for the first stage and second stage is low.

Despite the saliency maps after the first stage are not refined, salient object can be well detected after the second stage. Figure 2c and d show the visual comparison between the first stage and the second stage. This can be explained as follows. In terms of the spatial distribution, the salient pixels always gather together. And in terms of the feature distribution, the pixels in salient object inner are homogeneous in

appearance and highly connected to each other. While the pixels in background regions are the opposite.

4 Experimental results

This section evaluates the results of our approach. We evaluate our method on two public datasets. The first one is the MSRA-1000 (ASD) dataset, which contains 1,000 images provided by [11] with the corresponding accurate human-labeled binary masks for salient objects. Another is the SED1 [32], which contains 100 images. In order to measure the effectiveness of our methods, like many saliency detection models, we evaluate all methods through precision, recall and F-measure. Giving a saliency map with saliency value which is normalized [0, 255], a set of binary images can be obtained by varying the threshold from 0 to 255. As a result, the precision-recall curve is generated based on the ground truth mask. The F-measure is the overall performance of precision and recall, which can be measured as:

$$F_\beta = \frac{(1 + \beta^2)\,\text{Precision} \times \text{Recall}}{\beta^2\,\text{Precision} + \text{Recall}}, \qquad (9)$$

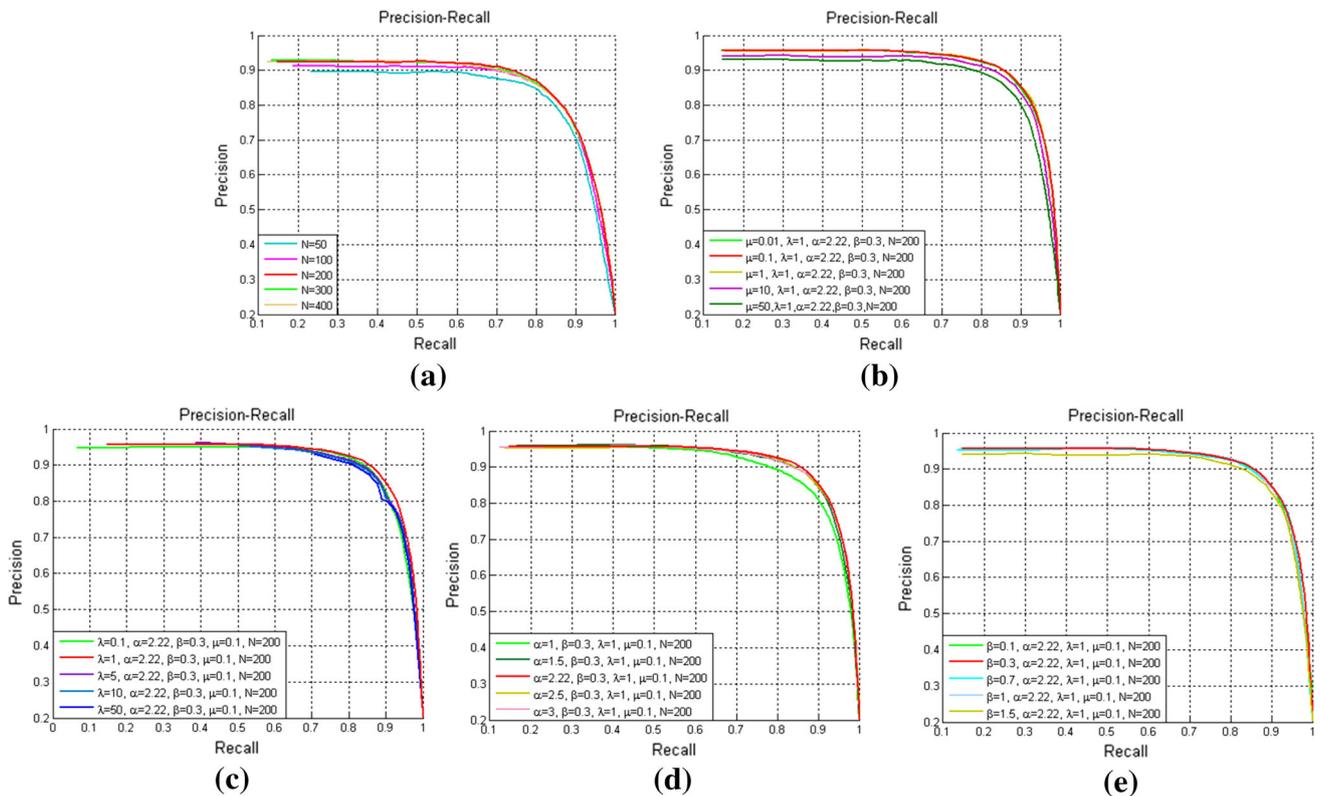where $\beta^2 = 0.3$ is according to [11].

Fig. 6 Precision-recall curves of different parameter settings on MSRA-1000 dataset. **a** The precision-recall comparison of different N on the coarse saliency maps. **b** The precision-recall comparison of different μ on the refined saliency maps. **c** The precision-recall compar-ison of different λ on the refined saliency maps. **d** The precision-recall comparison of different α on the refined saliency maps. **e** The precision-recall comparison of different β on the refined saliency maps

In our method implementation, a set of quantitative results comparison have been made by setting different parameter values in order to guide the selections of the parameters. We set the number of super-pixels $N = 200$, $\sigma^2 = \sigma_1^2 = 0.1$ in Eq. (2) and (6) respectively, $\mu = 0.1$ and $\lambda = 1$ in Eq. (5), and in Eq. (7), $\alpha = 2.22$, $\beta = 0.3$ for all the test images. Figure 6 shows the sensitivity of the proposed method to the parameters settings. As shown in Fig. 6a, the detection results are closely related to the scale of the image boundary super-pixels. In other words, the sizes of super-pixels affect the performance. When the number of super-pixels is small, each super-pixel has a big scale, some super-pixels may include the object regions, while if the number of super-pixels is too big, the scale of each super-pixel will be small (see Fig. 7), and the boundary information will not contribute much to the detection algorithm. These two situations are adverse to the detection effect. We also note that precision-recall curves are similar when $N = 200, 300, 400$. Therefore, considering the
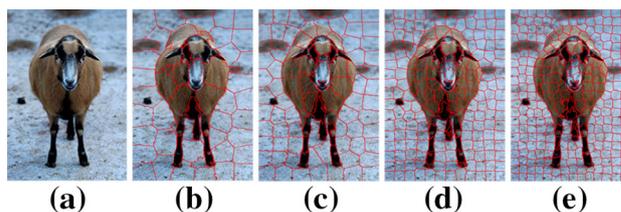

Fig. 7 Illustration of super-pixel maps. **a** Input image. **b–e** The maps with different super-pixel number $N = 50$, $N = 100$, $N = 200$, $N = 300$. The map with $N = 50,\ 100$ can not preserve the object boundary well, i.e., some super-pixels may include the object regions

computational complexity, we select the parameter $N = 200$ for all experiments.

4.1 Comparison with other methods

We compare our method with 16 prior saliency detection algorithms on MSRA-1000 dataset, including IT [3], AC
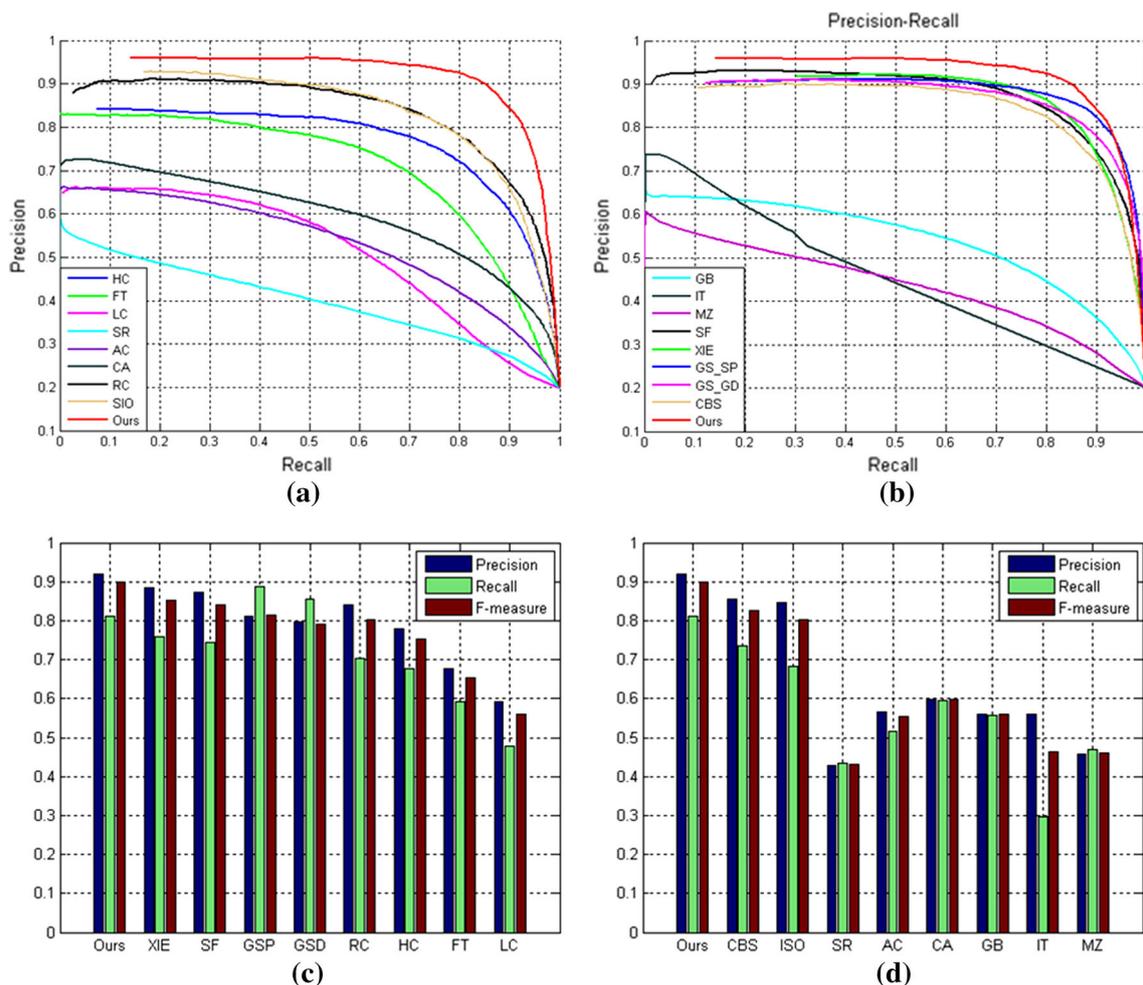

Fig. 8 Statistical comparison on MSRA-1000 dataset. **a, b** The precision-recall curves of different methods. **c, d** The precision, recall and F-measure for adaptive threshold
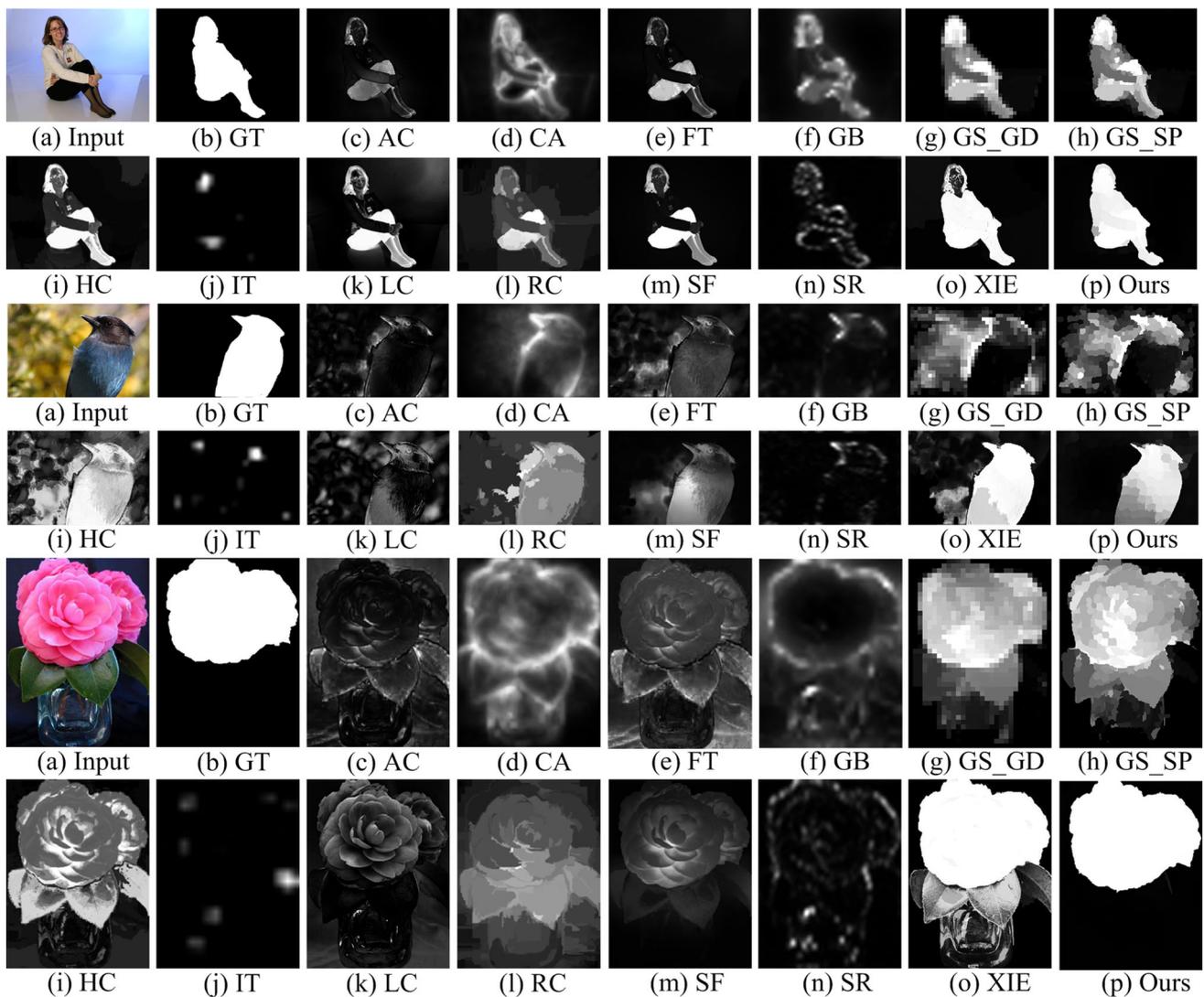
Fig. 9 Visual comparison of the saliency maps on MSRA-1000. From the above comparison, we can see that our method consistently generated better saliency maps

[10], FT [11], RC [13], HC [13], CA [14], GB [16], SR [17], MZ [20], SF [21], LC [23], XIE [27], GS_SD [28], GS_SP [28], CBS [33] and ISO [34]. In order to make a fair evaluation, we obtain the saliency maps of RC, HC, FT, LC, SR, AC, CA, GB, IT and MZ from [13]. For GS_SD, GS_SP, and SF, we directly use author-provided saliency results. For XIE, CBS and ISO, we run the authors' code. The precision-recall curve and F-measure are shown in Fig. 8a–d respectively, which indicate the different saliency maps emphasizing the effect of saliency, and providing a fair evaluation of different algorithms. The visual comparison results are shown in Fig. 9.

We compare our method with five classic saliency models on SED1dataset: FT [11], RC [13], HC [13], SR [17] and LC [23]. We obtain the saliency maps by using the code from the project page of Cheng [13]. Figure 10 highlights the visual comparison about different algorithms. The precision-recall curve and F-measure are shown in Fig. 11, which illustrate that the SED1 dataset is more challenging.

From the experiment results on MSRA-1000 and SED1 dataset, either the visual comparison or the quantitative comparison that is demonstrated by precision-recall curve and F-measure can indicate that our method significantly outperforms other classical methods in saliency detection. In addition, our algorithm can also achieve good effect in cluttered background.

4.2 Evaluation of color contrast based boundary information

As discussed in Sect. 2, we define the saliency of a super-pixel as the color relevance to the super-pixels on the four boundaries of image respectively. In order to demonstrate the
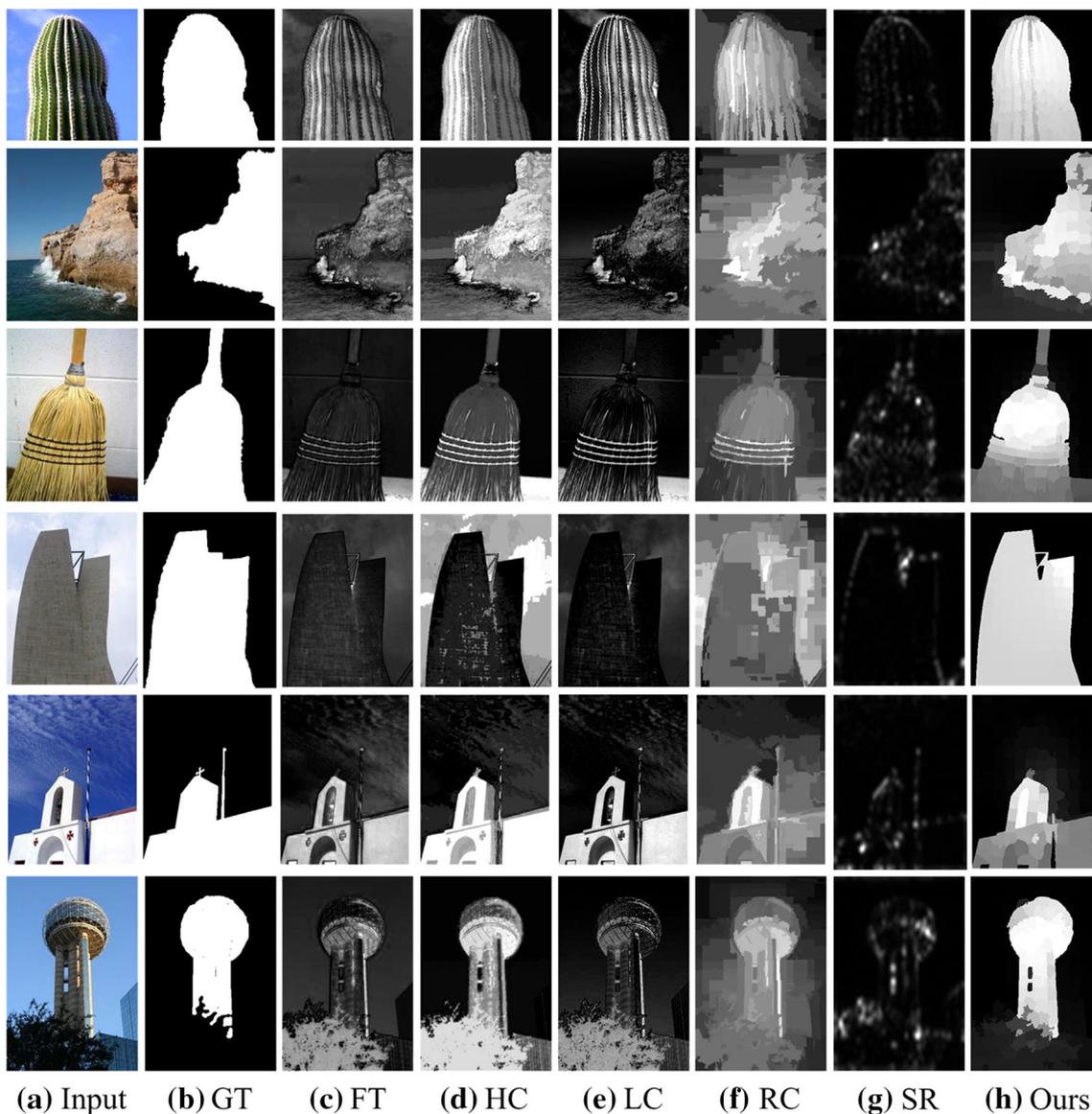
|**(a)** Input | **(b)** GT | **(c)** FT | **(d)** HC | **(e)** LC | **(f)** RC | **(g)** SR | **(h)** Ours |

Fig. 10 Visual comparison of the saliency maps on SED1 dataset. The comparison shows that our methods can achieve a better performance

effect of the proposed method, we compute two precision-recall curves for the coarse saliency maps with different measures. The first is to process the four boundaries of image integrally. The resulting curves are shown in Fig. 12a. Note that, the curve of processing integrally the four boundaries of image is slightly lower. That is because processing the four boundaries of image respectively suppresses the background more effectively. The second is the comparison of our coarse saliency maps with others which have been obtained from the previous model such as the AC [10], FT [11], RC [13], HC [13], CA [14], SR [17] and LC [23], by the precision-recall curve. Figure 12b shows the statistical comparison on MSRA-1000 dataset, where the red line represents the proposed coarse saliency map. The result shows that our method

has a better performance than most of other models. Figure 2c and 4c show the visual results about the coarse saliency maps.

4.3 Validation of the new energy function method

Based on the coarse saliency map and local smoothness properties, we propose a new energy function that takes the coarse saliency map as input. In order to evaluate the property of the new energy function, we take a simple comparison between the first and second stage through the precision-recall curve. Figure 12c shows the comparison result, where the red line represents the second stage result and the green line refers to the first stage result. The result indicates that by processing
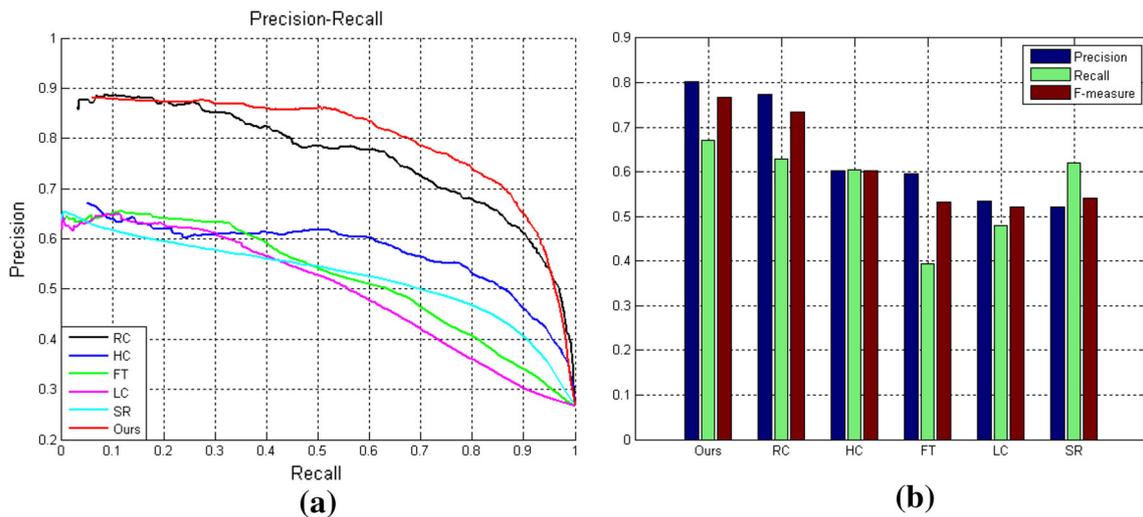
Fig. 11 Quantitative comparison on dataset SED1. **a** The PR curve about our algorithm and other five algorithms. **b** F-measure shows that although our algorithm does not have a high value, it is better than other methods
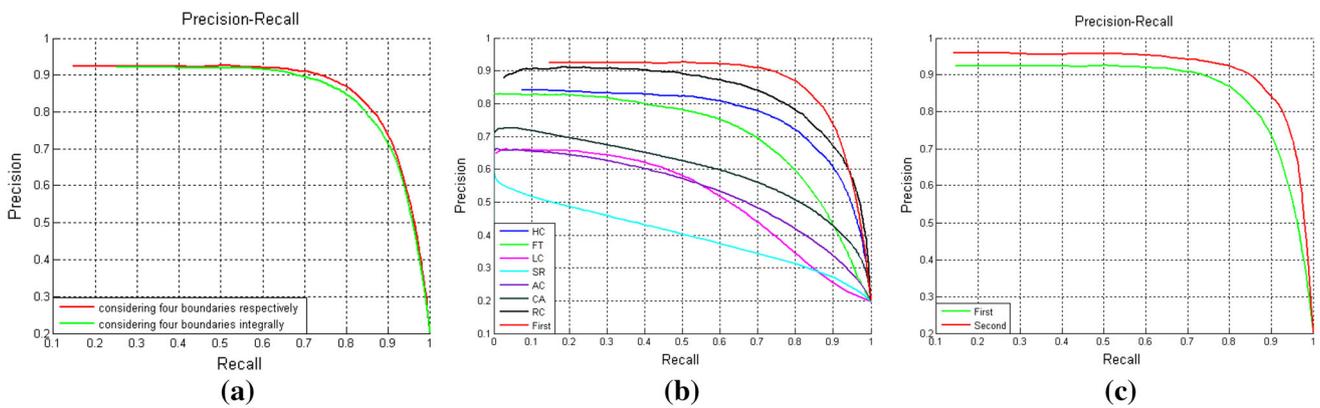

Fig. 12 Precision-recall curves to measure the effectiveness of our algorithm on MSRA-1000 dataset. **a** The comparison between considering the four boundaries of image integrally and respectively. The precision-recall curve in **b** is the comparison of the coarse saliency maps (the first stage) and the previous saliency models. **c** The precision-recall curve comparison between the first stage saliency maps and the refined saliency maps on the second stage

with the energy function, the coarse saliency maps achieve performance improvements. Figure 2c and d show the first stage and the second stage visual results respectively. It is obvious that the saliency maps through the new energy function have a better appearance and the salient object can highlight uniformly and the background can be suppressed effectively.

5 Conclusion

This paper presented a novel bottom-up saliency model based on the color contrast, and a new energy function which can generate refined saliency map according to the coarse saliency map is proposed. Saliency detection is carried out in a two-stage scheme to generate the saliency maps. We
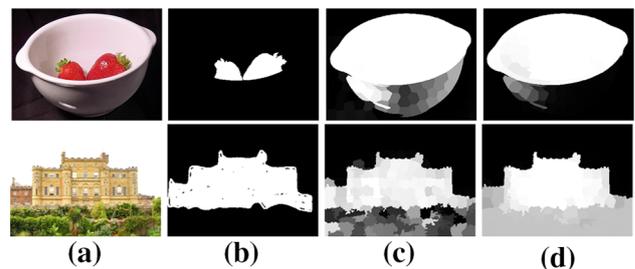

Fig. 13 The failure cases of the proposed method. **a** Input images. **b** Ground truth. **c** The first stage saliency maps. **d** The second stage saliency maps

evaluate the proposed algorithm on two public datasets and demonstrate the highly efficiency of our method that performs much better than those state-of-the-art methods.

As our saliency method is a two-stage method and the second stage is closely related to the first stage, if the saliency map that obtained in first stage is incorrect in most position of the saliency objects, the final saliency map would not have a good appearance. Figure 13 shows this failure case.

In the future, we plan to investigate efficient methods that incorporate more features of image to achieve a better performance and display low computational complexity.

References

1. Yarbus, A.: Eye movements and vision. Plenum Press, New York (1967)
2. Neisser, U.: Cognitive psychology. Appleton-Century-Crofts (1967)
3. 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)
4. Tsotsos, J.: What roles can attention play in recognition? In: Proceedings of seventh IEEE international conference development and learning, pp. 55–60 (2008)
5. Privitera, C.M., Stark, L.W.: Algorithms for defining visual regionsof-interest: comparison with eye fixations. IEEE Trans. Pattern Anal. Mach. Intell. **22**(9), 970–982 (2000)
6. Siagian, C., Itti, L.: Rapid biologically-inspired scene classification using features shared with visual attention. IEEE Trans. Pattern Anal. Mach. Intell. **29**(2), 300–312 (2007)
7. Gao, D., Han, S., Vasconcelos, N.: Discriminant saliency, the detection of suspicious coincidences, and applications to visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. **31**(6), 989–1005 (2009)
8. Guo, C., Zhang, L.: A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. IEEE Trans. Image Process. **19**(1), 185–198 (2010)
9. Itti, L.: Automatic foveation for video compression using a neurobiological model of visual attention. IEEE Trans. Image Process. **13**(10), 1304–1318 (2004)
10. Achanta, R., Estrada, F., Wils, P., et al.: Salient region detection and segmentation. In ICVS (2008)
11. Achanta, R., Hemami, S., Estrada, F., et al.: Frequency-tuned salient region detection. In CVPR (2009)
12. Bruce, N., Tsotsos, J.: Saliency based on information maximization. In NIPS (2005)
13. Cheng, M.M., Zhang, G.X., Mitra, N.J., et al.: Global contrast based salient region detection. In CVPR (2011)
14. Goferman, S., Zelnik-Manor, L., Tal, A.: Context-aware saliency detection, In CVPR (2010)
15. Gopalakrishnan, V., Hu, Y., Rajan, D.: Random walks on graphs for salient object detection in images. IEEE TIP (2010)
16. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In NIPS (2006)
17. Hou, X., Zhang, L.: Saliency detection: a spectral residual approach. In CVPR (2007)
18. Lu, Y., Zhang, W., Lu, H., et al.: Salient object detection using concavity context. In ICCV (2011)
19. Wang, L., Xue, J., Zheng, N., et al.: Automatic salient object extraction with contextual cue. In ICCV (2011)
20. Ma, Y., Zhang, H.: Contrast-based image attention analysis by using fuzzy growing. ACM Multimedia (2003)
21. Perazzi, F., Krahenbuhl, P., Pritch, Y., et al.: Saliency filters: contrast based filtering for salient region detection. In CVPR (2012)
22. Wang, W., Wang, Y., Huang, Q., et al.: Measuring visual saliency by site entropy rate. In CVPR (2010)
23. Zhai, Y., Shah, M.: Visual attention detection in video sequences using spatiotemporal cues. ACM Multimedia (2006)
24. Liu, T., Yuan, Z., Sun, J., et al.: Learning to detect a salient object. IEEE PAMI (2011)
25. Yang, J., Yang, M.: Top-down visual saliency via joint crf and dictionary learning. In CVPR (2012)
26. Li, J., Levine, M.D., An, X., et al.: Visual saliency based on scale-space analysis in the frequency domain. IEEE Trans. Pattern Anal. Mach. Intell. **35**(4), 996–1010 (2013)
27. Xie, Y.L., Lu, H.C., Yang, M.H.: Bayesian saliency via low and mid level cues. IEEE TIP **1**, 6 (2013)
28. Wei, Y.C., Wen, F., Zhu, W.J., et al.: Geodesic saliency using background priors. In ECCV **2**, 6 (2012)
29. Achanta, R., Smith, K., Lucchi, A., et al.: Slic super-pixels. Technical report, EPFL, Tech. Rep. 149300 (2010)
30. Levin, A., Lischinski, D., Weiss, Y.: A closed-form solution to natural image matting. IEEE Trans. Pattern Anal. Mach. Intell. **30**(2), 228–242 (2008)
31. Tatler, B.: The central fixation bias in scene viewing: selecting an optimal viewing position ndependently of motor biases and image feature distributions. J. Vis. **7**(14), 1–17 (2007)
32. Alpert, S., Galun, M., Basri, R., Brandt, A.: Image segmentation by probabilistic bottom-up aggregation and cueintegration, In CVPR (2007)
33. Jiang, H., Wang, J., Yuan, Z., et al.: Automatic salient object segmentation based on context and shapeprior. In BMVC (2011)
34. Li, J., Tian, Y., Duan, L., et al.: Estimating Visual saliency through single image optimization. IEEE Signal Process. Lett. **20** (2013)

Min Xu received the B.E. degree in communication engineering from Hunan Institute of Science and Technology, Yueyang, China, and the M.S. candidate student in Information science and Engineering at Hunan University, Changsha, China. His current research interests include saliency detection and feature extraction.

Hanling Zhang is an associated Professor at the Department of Information Sciences and Engineering at Hunan University since 2003. He has a PhD in Signal and Information Process from Northwestern Polytechnical University, China, 2003. From 2007 till 2008 he was a visiting professor at University of western Ontario, Canada. His research interests include machine learning, image processing.