Salient object detection based on regions

Article *in* Multimedia Tools and Applications · February 2014

DOI: 10.1007/s11042-012-1040-1

Xiaocong Zhou
Sun Yat-Sen University

Multimed Tools Appl (2014) 68:517–544
DOI 10.1007/s11042-012-1040-1

Salient object detection based on regions

**Zhuojia Liang · Mingjia Wang · Xiaocong Zhou ·
Liang Lin · Wenjun Li**

Published online: 25 March 2012

Abstract Salient object detection aims to automatically localize the attractive objects with respect to surrounding background in an image. It can be applied to image browsing, image cropping, image compression, content-based image retrieval, and etc. In the literature, the low-level (pixel-based) features (e.g., color and gradient) were usually adopted for modeling and computing visual attention; these methods are straightforward and efficient but limited by performance, due to losing global organization and inference. Some recent works attempt to use the region-based features but often lead to incomplete object detection. In this paper, we propose an efficient approach of salient object detection using region-based representation, in which two novel region-based features are extracted for proposing salient map and the salient object are localized with a region growing algorithm. Its brief procedure includes: 1) image segmentation to get disjoint regions with characteristic consistency; 2) region clustering; 3) computation of the region-based center-surround feature and color-distribution feature; 4) combination of the two features to propose the saliency map; 5) region growing for detecting salient object. In the experiments, we evaluate our method with the public dataset provided by Microsoft Research Asia. The experimental results show that the new approach outperforms other four state-of-the-arts methods with regard to precision, recall and *F*-measure.

Keywords Salient object detection · Saliency features · Center-surround · Color-distribution · Region growing and combination

Supported by National Natural Science Foundation of China (under Grant nos.60970156 and 61173082), Fundamental Research Funds for the Central Universities (Grant nos.2010620003162041 and 10LGZD05), and SYSU-Sugon high performance computing typical application projects (Grant no. 62000-1132001).

Z. Liang · M. Wang · L. Lin · W. Li (✉)
School of Software, Sun Yat-sen University, Guangzhou, China
e-mail: lnslwj@mail.sysu.edu.cn

Z. Liang · M. Wang · X. Zhou
School of Information Science and Technology, Sun Yat-sen University, Guangzhou, China

1 Introduction

Studies on visual attention [16] show that given an image the human brain and nervous system usually more concerns about certain objects, referred as salient objects. Salient objects detection is an important research topic and can be applied to image retrieval [8, 9], image compression [24], image browsing [19, 20], image cropping [27], and many other fields [13, 14, 32].

In general, salient object detection consists of two steps: the computation of the saliency map, and the output of the salient object. The saliency map indicates the saliency of each pixel in an image. Based on the saliency map, some filtering rules or selection strategies are applied to extract the salient object.

Early researches on salient object detection got the saliency map from the visual attention model directly. Itti and Koch [6] proposed a visual attention model, which was the realization and extension of a second biologically plausible architecture built by Koch and Ullman [10]. The low level characteristics of an image making a pixel salient are discussed in [22, 23, 31]. In 2001, Itti and Koch surveyed on the bottom-up model of visual attention [5], and illustrated five trends of the visual attention model. On the contrary, top-down approaches [3, 7, 28] are object-oriented and make use of prior knowledge about the scene or the context to identify the salient regions. The integration of top-down and bottom-up approaches were discussed in [15, 21, 35].

Ma and Zhang proposed a feasible and fast approach [18] to attention area detection based on contrast analysis. By analyzing the log-spectrum of an input image, the method in [4] extracted the spectral residual of an image in spectral domain, and proposed a fast method to construct the corresponding saliency map in spatial domain. In [16], Tie etc. proposed a set of novel features including multi-scale contrast, center-surround histogram, and color spatial distribution to describe a salient object locally, regionally and globally. A Conditional Random Field is learnt to effectively combine these features for salient object detection. Gopalakrish nan etc. presented a robust salient region detection framework based on the color and orientation distribution in an image [33].

These afore mentioned approaches for salient object detection adopt low-level (i.e. pixel-based) features, whereas the following are based on region or object representations. Thomas et al. proposed a region-oriented visual attention framework [29]. In [30], the saliency map with region information from image segmentation was further enhanced and the most salient region (proto-object) was selected. Then, regions were organized using perceptual groupings, and their pop-out sequence was determined. On the basis of [16], Zhuang proposed a new salient object detection algorithm based on segments [11].

Although existing salient object detection approaches perform well in many cases, they fail to find out the complete salient object in some situations due to the intrinsic weakness in these approaches. These methods are lack of global organization and inference which is an important feature in human visual system. For human visual perception, homogeneous regions should have the same saliency, but the saliency maps may not. For example, the salient object in Fig. 1(a) should be the whole red leaf, so the leaf should have the same saliency. But it is easy to observe from the saliency image in Fig. 1(c), (saliency is represented by the intensity) that the lower right corner is less salient than other part of the leaf. This is the general drawback of many pixel-based methods. Recently a few region-based salient object detection approaches were proposed to overcome this kind of weakness. But these approaches led to similar results as the pixel-based methods because they also are lack of global organization and inference.

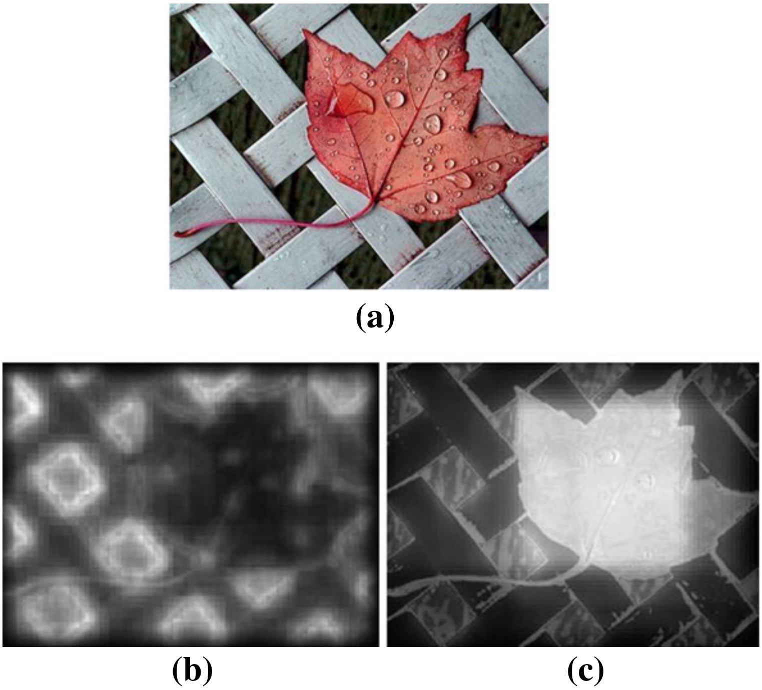


Fig. 1 **a** Original image. **b** Failed saliency image from [6]. **c** Saliency image from [16]

So we need global grouping and inference as human visual system. We can compute the saliency based on regions and let regions become the basic computing units. Furthermore, a salient object may consist of one or more homogeneous region(s). While processing the output of the salient object, we can restore the whole object by merging the homogeneous regions. Therefore, the salient object detection algorithm makes use of homogeneous regions as the basic computing units. Region-based salient object detection can solve the problem that the pixels in the homogeneous regions have different saliency. At the output processing, it can restore the salient object more accurately and improve efficiency.

Some salient object detection methods(as [16, 29]) only compare with the neighboring regions when computing the region saliency. The resulting regional features then cause the saliency computation sink into the local regions. Due to the lack of high level or global features, regions with high saliency are found out to be obvious background regions. As shown in Fig. 1(b), the leaf fails to be the salient object. The failure in this case indicates that not only local features but also global features are needed.

From above, we need more effective region-based features. So two basic saliency features are proposed in this paper: area and position. A saliency survey is conducted to justify that the area and position of a region hold an important impact on saliency. The survey also quantifies the impact by the probability distribution of the area and position.

The remainder of this paper is organized as follows. The area and position saliency features are introduced in section 2. Then the framework of our approach is described in Section 3. The experimental results in Section 4 show that our approach is better than other four existing approaches. Section 5 concludes the paper and discusses our future work.

2 Saliency features: area and position

Most of saliency computation approaches concentrate on the analysis of color, brightness and orientation. In fact, these features can affect the saliency of the pixels. However, if the computation of saliency is based solely on these features, the performance and efficiency will not be satisfactory, because they are the low-level and pixel-based features.

To take the characteristics of regions into consideration, two novel saliency features are proposed in this paper: area and position. The area is the minimum bounding rectangle of the region, and the position is the center point of the area. The area and position saliency features are regional characteristics. The combination of these two features with other low-level features can result in a better detection of salient objects.

2.1 Saliency survey for area and position

If an object in the image occupies a small region, it will not be easily to attract human attention. If an object in the image occupies a very large region and distributes almost in each part of the image, human would tend to think this object as a background and not as saliency. An object with suitable area is more salient than the others. In addition, human will generally focus on the central part of their vision, and tend to consider an object in the center as the most important. That is why they like to put the important objects at the central place when taking a photo. Therefore, an object at the center of the image is more likely to be considered as important and salient than the others.

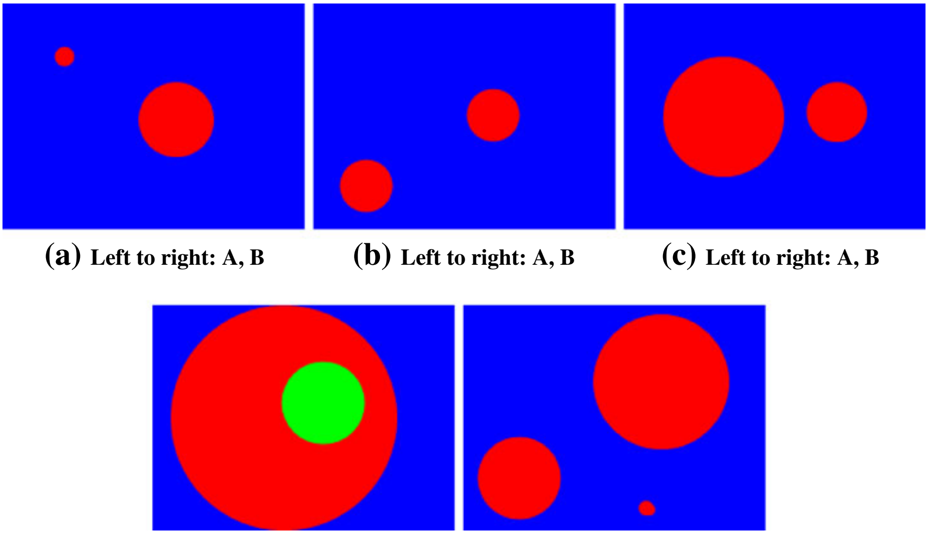
In a word, the area and position affect the saliency of the object. If the object occupies a suitable area (neither too small nor too large) and it is closer to the center of the image, it is more possible for this object to become the salient object. In order to justify this claim, a saliency survey is conducted as following.

Five pictures are designed, as shown in Fig. 2, to test how the area and position of a region affect the saliency. There are some circles in each picture. The circles in a single picture have different areas and positions. The questionnaire survey involved 59 subjects. We sent emails to subjects. The subjects gave the order of circles according to saliency in every picture. If they thought they had the same saliency, they could give them a bracket. We got results and summarized. The subjects under test are asked to tell which circle is the most salient. In order to avoid the impacts caused by the color, brightness and shape, most of these pictures are solely in red, green and blue, and all the objects in a single picture are circles. The results shown in Table 1 indicate that the area and position of the circle have a strong impact on the saliency of the circle. It also concludes that an object tends to be more salient if the size of the object is suitable and the position of the object is closer to the center of the image.

This conclusion can be utilized to compute the saliency. But quantitative measurement of the impacts of area and position need to be studied, for example, what is the suitable area? How is near the center of an image? The remaining parts of this section will answer these two questions.

2.2 Area as a saliency feature

In order to make the property of area directly guide the computation of saliency, the impact of area on saliency need to be quantified. To answer the question how the impact is estimated, we analyze the area of real world salient objects in the MSRA salient object



(a) **Left to right: A, B** **(b)** **Left to right: A, B** **(c)** **Left to right: A, B**

(d) **Red: A, green: B** **(e)** **Biggest: A, middle: B, smallest: C**

Fig. 2 Pictures for saliency survey

image set provided by Microsoft Research Asia.[1] The subset *Set B* in this image set contains 5,000 images. The salient object of each image is tagged by 9 minimum bounding rectangles, which are drawn by 9 different persons. Because the images in *Set B* come from wide variety of sources, not from a certain field, it is a general and representative data set and the experiment based on this data set is also representative.

We randomly select 1,000 images from *Set B* to compute the probability of the area ratio of the salient object tagged rectangles. The test is performed four times, and the probability histograms are shown in Fig. 3, where the x-coordinate indicates the area ratio of the salient object and the image area, and the y-coordinate shows the value of probability. The high similarity of the four results reveals that the probability distribution of the area ratios is general for different kinds of pictures.

After analyzing the results of the four tests, we can find out that the maximal probability is at the point where area ratio is 0.35, and slowly descends to the left and the right sides. This probability distribution matches the previous saliency survey well. The area ratio of 0.35 corresponds to the aforementioned suitable area. Furthermore, the area ratio from 0.2 to 0.6 meets the condition of the area not too small or too large. The relationship between the area ratio and saliency is thus revealed by analyzing the probability distribution of the area ratio. Then the probability of the area ratio can be utilized to estimate the impacts of the area on saliency.

As a result, the area is proposed as a saliency feature like color, texture and orientation. To establish a formula to help compute the area saliency feature in this paper, we randomly select another 1,000 images from *Set B* to perform the area ratio probability test once more, and then smooth the probability histogram, as shown in Fig. 4. The purpose of smoothing is to reduce the gap between two adjacent regions to enhance the robustness. The smooth

[1] http://research.microsoft.com/en-us/um/people/jiansun/salientobject/salient_object.htm

Table 1 How area and position have impacts on saliency

Properties of Circles in Testing Pictures				Results
		Area	Position	
(a)	A	too small	away from center	3%
	B	suitable	near center	97%
(b)	A	suitable	away from center	22%
	B	suitable	near center	58%
(c)	A	suitable	near center	93%
	B	small	near center	3%
(d)	A	too big	near center	17%
	B	suitable	near center	80%
(e)	A	suitable	near center	81%
	B	small	far from center	17%
	C	too small	far from center	2%

For picture (b),(c),(d),the sum of results is not 1, because some testers thought they had the same saliency. In Table 1, we did not show the results of the same saliency

probability histogram can be expressed by Formula 1, where x is the area ratio, and y is the value of the area saliency feature, i.e. the probability.

$$y = areaSaliency(x) \tag{1}$$

2.3 Position as a saliency feature

According to the results of the saliency survey, an object tends to be more salient if it is closer to the center of the image. Just as the area, we need to quantify the impact of position on saliency.

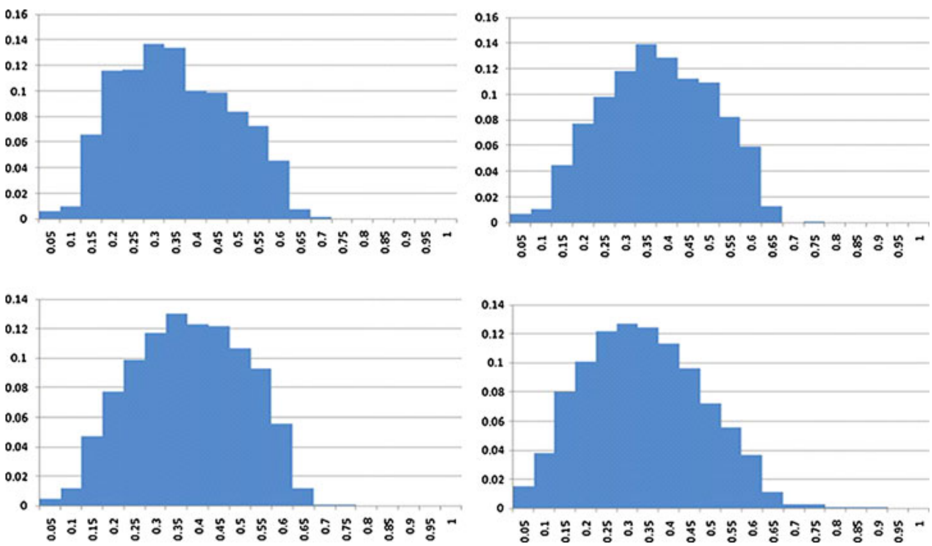


Fig. 3 Probability histograms of area ratios

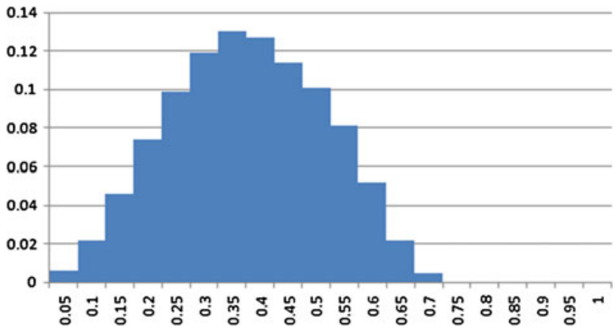


Fig. 4 A smooth probability histogram of area ratios

Experiments similar to the above section are conducted. 1,000 images are randomly selected from *Set B* to compute the probabilities of positions. Positions are located based on the center of the minimum bounding rectangle of the salient object in the image. After finding all the centers of salient objects (indicated by black points shown in Fig. 5(a)), we partition the whole image area into 20 regions (red rectangles shown in Fig. 5(b)).

Four probability histograms are generated from the experiments, as shown in Fig. 6, where x-coordinate represents the region indices following Fig. 5(b), and y-coordinate shows the probabilities. It can be concluded from the results that salient objects have higher probabilities to appear closer to the center of the image, the same result as the saliency survey. So it is reasonable to use the probability distribution of positions to estimate the impact of position on saliency.

Following the same approach to area saliency, we perform the test again to get the probability histogram, then smooth it to get the probability histogram as shown in Fig. 7. Because the position of the object has strong impact on saliency, position is proposed in this paper as another saliency feature. The smooth probability histogram in Fig. 7 can be expressed by Eq. 2, which is used to compute the position saliency feature.

$$y = positionSaliency(x) \tag{2}$$

In Formula 2, $x \in \{0, 1, \ldots, 19\}$ is the region index as shown in Fig. 5(b), and y is the value of the position saliency feature, i.e. the probability.

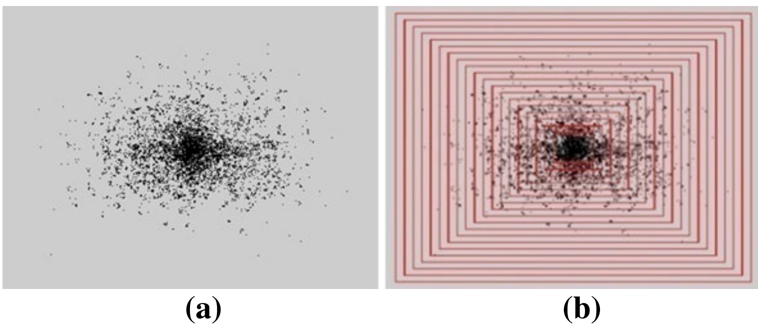


Fig. 5 **a** Centers of salient objects. **b** Region partition

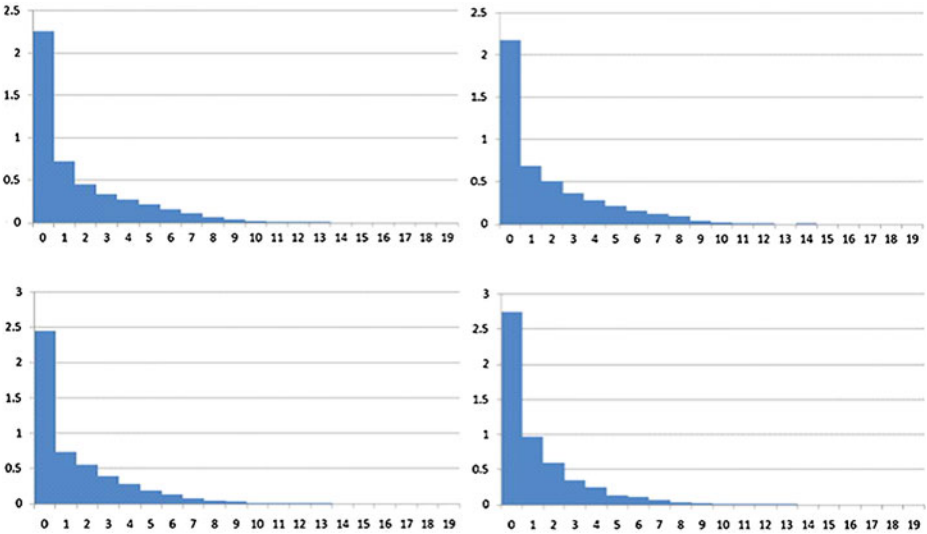


Fig. 6 Probability histogram of positions

3 The framework

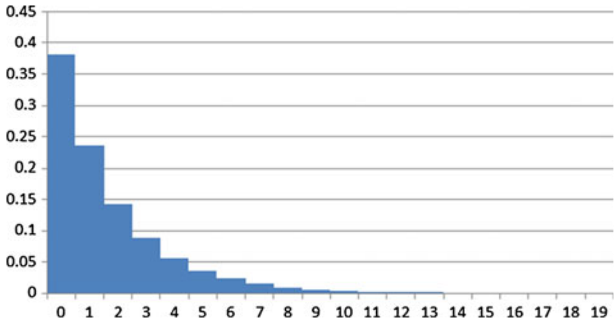
A region-based salient object detection approach is also proposed in this paper. The first step of this approach is to partition an image into several disjoint and homogeneous regions. Secondly, the homogeneous regions are clustered. And then the center-surround and color-distribution features are computed and combined to obtain the final saliency map. Finally, the salient object is found out after region growing and selection. The flowchart of the region-based salient object detection is shown in Fig. 8.

3.1 Image segmentation and region clustering

3.1.1 Image segmentation

Image segmentation is the process of partitioning an image into disjoint and homogeneous regions [17]. Regions of an image should be uniform and homogeneous with respect to some characteristics such as gray tone or texture. Segmentation is an important operation in many

Fig. 7 A smooth probability histogram of positions



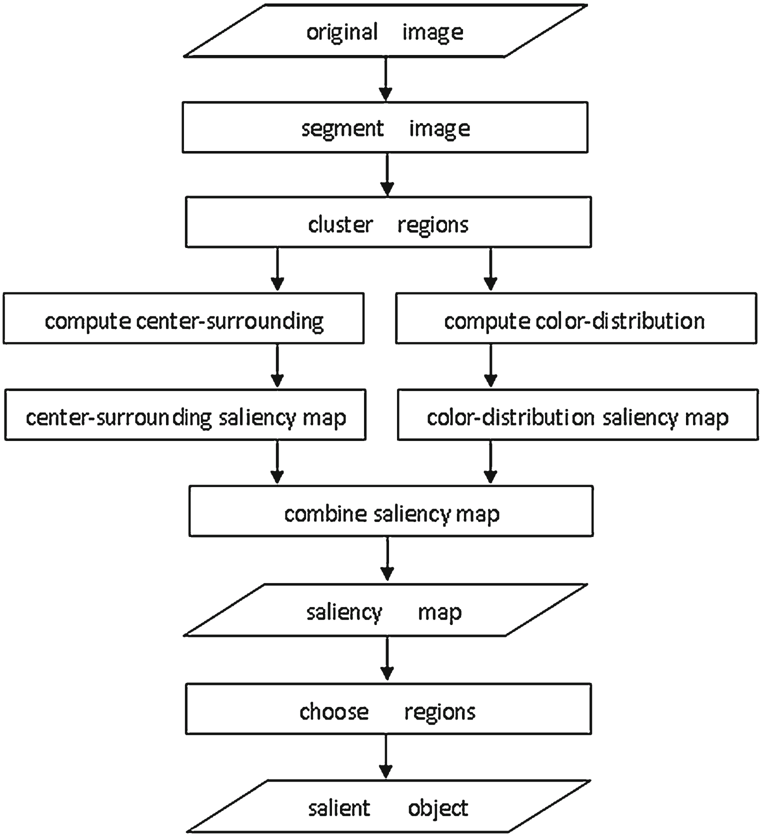


Fig. 8 Flowchart of salient object detection based on regions

applications of image processing and computer vision, since it is the first step of low-level processing of images.

Homogeneity is largely related to the local information extracted from an image and reflects how uniform a region is [1].Image segmentation needs a certain homogeneity predicate H. The segmentation of an image I is defined as a partition of I into a set of N regions R_n (n=1, …,N), satisfying that $H(R_n)$ = **true** for any n, and H $(R_n \cup R_m)$ = **false** for any adjacent R_n and R_m. It states that two adjacent regions cannot be merged into a single region that satisfies the predicate H. This H can be further used in region clustering.

Human visual attention is more related to region than pixel. So we can compute the saliency map based on regions instead of pixels. The region-based method can overcome some shortcomings of the traditional pixel-based methods. For example, the region-based method can ensure that the pixels in the homogeneous regions have the same saliency. Furthermore, an object in the image is often composed of a number of homogeneous regions. Through the combination of regions, the salient object can be found more easily. In a word, regions can be the basic computing units of the salient object detection.

The partitioning of the image segmentation algorithm can be controlled by the homogeneity predicate H, which has a set of default parameters. Many image segmentation algorithms [2, 12, 25, 26] with their default settings of parameters can get the homogeneity regions that well match the human visual perception, we use state-of-the-art to get the homogeneity regions.

3.1.2 Region clustering

After image segmentation, some disjoint and homogeneous regions are generated. The distribution of homogeneous regions in an image can affect the saliency of the regions. Homogeneous regions are clustered in order to estimate their distribution.

The homogeneity predicate H in the image segmentation stage can be used in region clustering. Actually H can also be redesigned for practical needs. The region R_n and R_m ($m \neq n$) belong to the same cluster if and only if $H(R_n \cup R_m)$ = **true**. In this paper, the clustering algorithm is designed based on graph theory. At the beginning, each region is considered as a vertex. There exists no edges, so all vertices are isolated. For any two vertices R_n and R_m ($m \neq n$), $H(R_n \cup R_m)$ is computed; if the result is **true**, an edge will be added connecting R_n and R_m. Finally, each connected graph becomes a cluster.

For the original image in Fig. 1(a), image segmentation is generated by JSEG [2], as shown in Fig. 9(a), with each region surrounded by white lines. Figure 9(b) shows the region cluster for white stripes.

3.2 Saliency map

Traditional saliency maps record the saliency of each pixel in the image, expressed by gray intensity. The larger the gray value is, the more salient the pixel is considered. But the

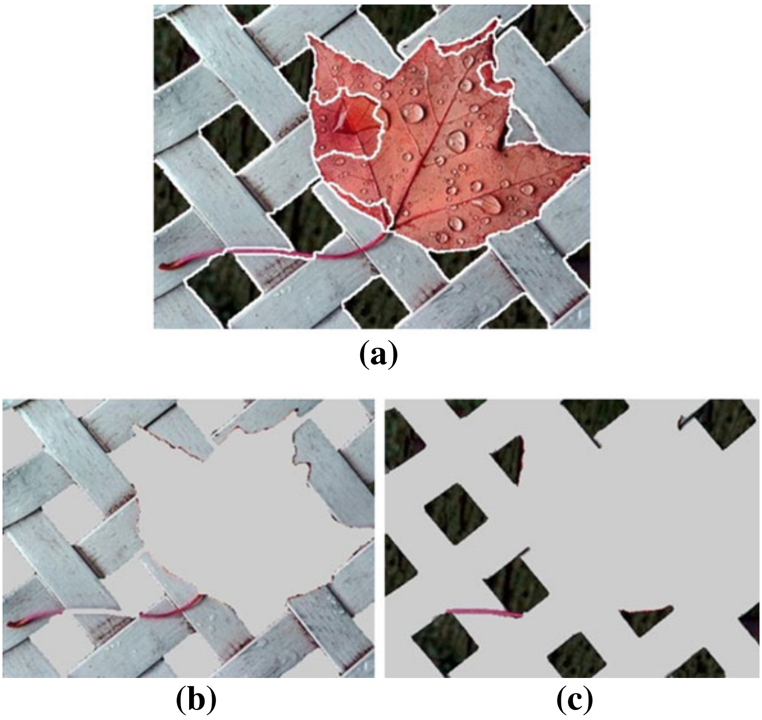


Fig. 9 **a** Image segmentation. **b**,**c** Region clustering

saliency map proposed in this paper is based on regions. It means that the pixels in the same region have the same saliency. Saliency map hereby is composed of the region-based center-surround feature and the color-distribution feature, which are achieved through the analysis of the three basic saliency features, including the color, position and area of the region.

3.2.1 Center-surround feature based on regions

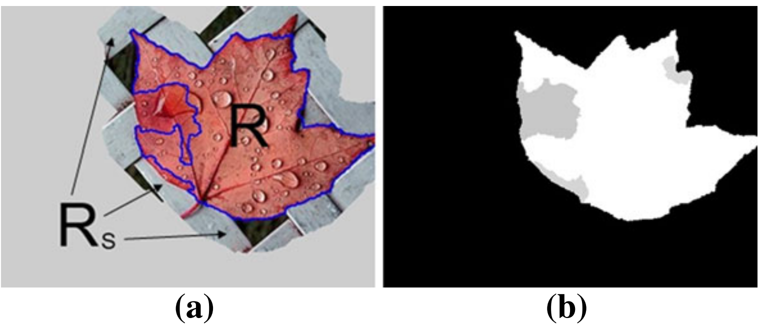
A center-surround feature is defined as the appearance variance of a certain region with respect to its surrounding regions R_s. The larger the variance, the more salient the region R. Since a region is compared only with its surrounding areas, this feature is a region-based feature. In existing approaches, the center-surround feature is computed at the pixel level, which gives rise to the problem of failure to fix the size, shape and position of region R.

For example, while figuring out the center-surround feature in [16], rectangles with different length-width ratio are utilized to express the size of the central region R and its surrounding region R_s. This method need compute many times for a center region, and the center region may be not in the range of different length-width ratios. This leads to low efficiency and weak performance. The central surrounding feature based on regions has advantages over pixel-based features.

The area within the blue lines in Fig. 10(a) is denoted as R. The surrounding region R_s is obtained through the expansion of the border of region R by certain thickness. The thickness of the expansion is set to 35 pixels in this paper, since this is the reasonable human vision area when comparing the color of different regions. The feature of a region is represented by the color histogram of the region. Given the color histogram of region R as $histogram_R$, and that of the surrounding area R_s as $histogram_{Rs}$, we can figure out the center-surround feature of region R with the commonly used distance formula as shown in Formula 3, where N is the dimension of the color histogram.

$$centerSurround_R = \sqrt{\sum_{i=0}^{N-1} \left(histogram_R[i] - histogram_{R_S}[i]\right)^2} \tag{3}$$

For every region R_j in the picture, we can compute the $center\ Surround_{Rj}$ for R_j. Then, the center-surround feature values of all regions are normalized to an interval between 0~255 to generate a grey image, where 0 is the minimum value and 255 is the maximum value respectively, and other value is set according to ratio. As a result, the center-surround feature image of Fig. 1(a) is shown in Fig. 10(b).



(a) **(b)**

Fig. 10 **a** Center-surround feature based on regions. **b** Center-surround feature image

A special case need to be further discussed. Given the region clustering of a certain region as C, if the minimum bounding rectangle of C is too large (for example, the area ratio is larger than 0.7), the resulting area is probably a background of the image. In order to reduce the computation and avoid disturbance for this case, we set the value as 0, rather than computing the central surrounding features for these regions. As the example shown in Fig. 9(c), although the black regions has strong contrast, the minimum bounding rectangle of region cluster is two large , we think it is background.

3.2.2 Color-distribution feature based on regions

Color-distribution feature is another common feature used to figure out saliency. It is a global feature describing the distribution of some certain color in an image. The more concentrated the color distributes, the more salient the regions with this color will be, and vice versa.

After the color-distribution features based on regions are computed hereby, the distribution of regions with the same feature in the image will be figured out. Because the area and position of a region indicate the distribution of that region in the whole image, we use the area and position saliency to compute the distribution feature of regional color. While area is the description of the size of an object, position is the description of the azimuth of the object in the image. These two features are independent from each other. Therefore, the impacts on saliency of a region made by the area and position can be regarded as independent events. Furthermore, the above area and position saliency features are computed on base of probability distribution, so the product of these two features can measure the whole impact made by the area with the position, according to the probability distribution theory.

Assume that a region R in the image I is in the region cluster C. The minimum bounding rectangle of region R is $MinRect_R$ and the region partition index of the center of $MinRect_R$ is N_R. The minimum bounding rectangle of the region cluster C is $MinRect_c$, and the region partition index of the center of $MinRect_c$ is N_c. The color-distribution feature of R is computed following Eqs. 4, 5 and 6.

$$colorDistribution_R = areaSaliency\left(\frac{area(Min\mathrm{Rect}_R)}{area(I)}\right) \times positionSaliency(N_R) \tag{4}$$

$$colorDistribution_C = areaSaliency\left(\frac{area(Min\mathrm{Rect}_C)}{area(I)}\right) \times positionSaliency(N_C) \tag{5}$$

$$colorDistribution_R = \min(colorDistribution_R, colorDistribution_C) \tag{6}$$

area() is the area function for the region of the minimum bounding rectangle. When Eq. 6 is applied to every region in the image, the respective color distribution feature will be figured out. The results will be normalized to the interval between 0~255 as we did with center-surround features. The resulting grey image is shown as Fig. 11.

3.2.3 Combined saliency map

Center-surround feature are shown to well describe the regional saliency and the color-distribution feature encodes the global saliency. The combination of these two features can

Fig. 11 Color-distribution feature image

fully express saliency features of an image. Since all the values of center-surround features and color distribution features are normalized to an interval between 0 and 255, a simple linear combination is applied to combine these two features, as shown in Formula 7.

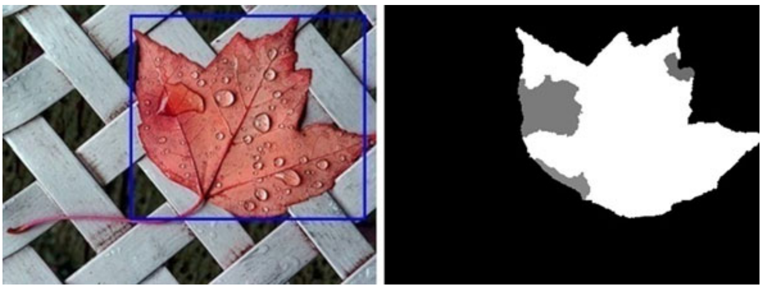
$$saliency_R = centerSurround_R + colorDistribution_R \tag{7}$$

Then the combined saliency value generated by Formula 7 is normalized back to an interval between 0 and 255, and leads to the final saliency image as shown in Fig. 12(b). Compared with the salient object drawn by MSRA *Set B* with the blue lines as shown in Fig. 12(a), the saliency image generated in this way tends to be more readable for human vision than that from traditional approaches (such as Fig. 1(b) and (c)).

3.3 Output of salient object

3.3.1 Aggregation

Usually the gravity center of an object with high aggregation locates inside the object, and the distances between the center and the borders of the object are relatively equal. In real world, most of objects appear with high aggregation. It is the same with the saliency object in images. This property can help find out the complete salient object from the saliency image. The property of aggregation refers to the measurement of intensiveness of an object in an image.



(a) **(b)**

Fig. 12 **a** Original image and salient object. **b** Saliency image obtained by our approach

The final result of salient object detection in this paper is a collection of clustered regions corresponding to the salient object in the image. Since a salient object usually appears with high aggregation, a region with low aggregation has high probability to be excluded from the composition of the salient object. This is the reason why we can guide the combination of regions by computing the aggregation of the region collection, and get a more complete and accurate salient object.

The aggregation of a region collection can be obtained by analyzing the superposition of regional minimum bounding rectangles. Assuming a region collection $C=\{R_1, R_2, \ldots, R_n\}$, where $R_i(1\leq i\leq n)$ denotes a region in C, and the minimum bounding rectangles of R_i and C are $MinRect_{Ri}$ and $MinRect_C$ respectively. The aggregation of region collection C is computed by Formula 8.

$$aggr_C = \frac{\sum_{i=1}^{n} area\left(Min\mathrm{Rect}_{R_i}\right)}{area\left(Min\mathrm{Rect}_C\right)} \tag{8}$$

As shown in Formula 8, the resulting $aggr_C \in (0, n)$. The larger the $aggr_C$, the higher the aggregation of region collection C. A region collection C is considered with high aggregation if $aggr_C > 1$, and vice versa. Experiences from our experiments reveal that the aggregation value of a region collection generally falls into the range of [0.3, 2].

3.3.2 Single salient object

Once the final saliency image is generated, the salient object of each region can be localized. The regional saliency value determines whether the region will be a candidate region for saliency object or not. Meanwhile, the set of candidate regions forming the complete salient object can be decided by the saliency feature and the aggregation of area and position. As shown in Fig. 13(a), when a balloon floats in the air, it is obviously the salient object in the image. Figure 13(b) and (c) show the saliency image and the resulting salient object respectively.

Salient Object Measure (SOM) When a region collection is obtained, we need to estimate the probability for the region collection to become the saliency object. Then the region collections with the highest probability become the final result. Assuming the minimum bounding rectangle of the region collection C in image I is $MinRect_C$, and the central

(a) (b) (c)

Fig. 13 **a** Image containing a single salient object. **b** Saliency image. **c** Salient object position of the rectangle is denoted as N_C, the Salient Object Measure (abbreviate to SOM) of C is computed following Formula 9.

$$SOM_C = areaSaliency\left(\frac{area(Min\mathrm{Rec}t_c)}{area(I)}\right) \times positionSaliency(N_c) \times aggr_c \qquad (9)$$

As shown in Formula 9, the value of SOM depends on both the area/position saliency feature and the aggregation feature. A larger SOM indicates the higher probability that the region collection becomes a component of a salient object. Considering area, position and aggregation are independent features; we make use of the product of these three features to combine their impacts on saliency.

Candidate regions and seed regions A candidate region is the region which is probable to compose a salient region, while a seed region is a special candidate region that can be used to guide the expansion or growth of the salient region collection. In order to get the candidate regions and the seed regions, the first step is to compute the average saliency of the saliency image, denoted as $saliency_{avg}$, which is the arithmetic mean of all pixels. Region R will not be considered as a part of the salient object if $saliency_R < saliency_{avg}$; if $saliency_R \geq saliency_{avg}$, R could be a candidate region; if $saliency_R \geq 2\times saliency_{avg}$,R can be the seed region; if there is no such region, the largest 3 saliency regions will be regarded as the seed regions. All of candidate regions and seed regions will become the parts of the salient object.

Regions growing and combination Assuming R_s is the seed region, the resulting saliency region collection is denoted as C_{sal} and initialized to the singleton $\{R_s\}$, and the processing region collection C_{adj} contains all candidate regions that are adjacent to any region in C_{sal}, then the region growing process is shown as follows.

(1) Compute the SOM of the current region collection C_{sal}, that is SOM_{Csal}.
(2) Sort descending the regions in C_{adj} according to SOM, and mark all regions in C_{adj} as untreated.
(3) Select the untreated region R in the sorting list with the highest SOM. If no such region exists, jump to step 6.
(4) Let C'be the new region collection containing R and all regions in C_{sal}, then compute its $SOM_{C'}$. If $SOM_{C'} \leq SOM_{Csal}$, mark the region as treated and jump to step 3.
(5) Remove R from C_{adj} and add it into C_{sal}, then add to C_{adj} all regions adjacent to R, and re-compute SOM_{Csal}. Jump back to step 2.
(6) Halt.

Each seed region will generate its respective region collection after the above growing process. The region collection with the largest SOM becomes the candidate region collection of the salient object.

3.3.3 Multiple salient objects

Sometimes several unconnected salient objects appear in a single image, as shown in Fig. 14 (a). If the aforementioned approach to the detection of single salient object is simply applied to the image in such images, some salient objects in the image will be missing. Therefore, the aforementioned approach needs to be tailored for the detection of multiple salient objects.

Fig. 14 **a** Image containing multiple salient objects. **b** Saliency image. **c** Salient objects

The detection of multiple salient objects is based on the results of the detection of single salient object. When dealing with the growing process of a single salient object, each seed region will be used for generating its respective region collection. Since the same region collection can be generated by two different seed regions, duplicate region collections will be discarded. It gives rise to only limited number of the remaining region collections generated by seed regions, usually no more than 4 regions. Then each of the combination of different region collections will be checked, based on the SOM of the combined region collections. The SOM is the product of the saliency features of the area and the position of the minimum bounding rectangle of the combined region collections.

Compared with the detection of single salient object, the detection of multiple salient objects does not consider the aggregation of nonadjacent region collections, and SOM depends solely on the product of the saliency features of area and position. Finally, the SOM of each single object and that of the combination of multiple objects are compared. The composite region of the salient object is the region collection with the largest SOM. As a result, the saliency image and the salient objects of Fig. 14(a) is shown as Fig. 14(b) and (c) respectively.

4 Experiments

In order to demonstrate the advantage of the proposed method, we design the following experiments: (1) empirical study of saliency features, (2) comparison with Color-Orientation Framework, (3) comparison with the state-of-the-arts approaches, and (4) analysis of robustness. At last we analysis the fail cases and the efficiency of the proposed approach.

4.1 Image dataset

Experiments are carried on with the *Microsoft Research Asia Salient Object Set B* [16], including 5,000 images tagged by 9 users. We use 1,000 images to train the parameters for the saliency feature of area and position, and the remaining 4,000 images for empirical evaluating.

4.2 Evaluation benchmark

The experiments are evaluated following the classical benchmark metrics, including *Precision*, *Recall* and *F-Measure*. The first step is to generate a *Saliency Probability Map* (SPM), following the detailed description in [16]. The SPM $G=(g_{ij} \mid g_{ij}\in [0,1])$ indicates the probability of the saliency for each pixel (i, j). It is computed from the 9 minimum bounding rectangle tagged in the image by 9 different persons. The result of salient object detection in this paper is a collection of regions. We use a binary mask $D=\{d_{ij}\}$ where $d_{ij}\in\{0, 1\}$ to represent this collection: the pixel d_{ij} is included if d_{ij}=1, otherwise d_{ij} is excluded. Given the image with height h and width w, the computation of *Precision*, *Recall* and *F-Measure* is shown in Formula 10, 11 and 12, where α is a positive real number deciding the importance of *Precision* over *Recall* while computing *F-Measure*. We set α = 0.3 following the same reason described in [33].

$$\mathrm{Pr}\,ecision = \frac{\sum_{i=0}^{h-1}\sum_{j=0}^{w-1} d_{ij} \times g_{ij}}{\sum_{i=0}^{h-1}\sum_{j=0}^{w-1} d_{ij}} \tag{10}$$

$$\mathrm{Rec}all = \frac{\sum_{i=0}^{h-1}\sum_{j=0}^{w-1} d_{ij} \times g_{ij}}{\sum_{i=0}^{h-1}\sum_{j=0}^{w-1} g_{ij}} \tag{11}$$

$$F - Measure = \frac{(1+\alpha) \times \mathrm{Pr}\,ecision \times \mathrm{Rec}all}{\alpha \times \mathrm{Pr}\,ecision + \mathrm{Rec}all} \tag{12}$$

4.3 Experiment 1: Different salient features

We propose the experiment to quantitatively validate different salient features and show the advantage of combining different salient features.

A. Center-Surround features

Center-Surround feature is region feature. To avoid noise (background) and reduce the number of salient objects, we think that if the area ratio of minimum rectangle of the region cluster is larger than 0.7, the region is background. This plays an important role in getting center-surround salient map. Figure 15 shows some results of center-surround salient map. The results are satisfied, such as in row (1, 2, 3). Our proposed method can also detect decentralized salient objects, as in row (5, 6). For complex scene, it works also well.

B. Color-Distribution salience map

We use the area and position feature to calculate color-distribution salience map. A few results are shown in Fig. 16. Some of the results look satisfied due to the effectiveness of the global feature, such as row (1), row (2), and row (4), but the feature work not well in the others examples.

C. Feature combination.

The center-surround feature is a regional feature, while the color-distribution feature is a global feature. They all have limitations. For the center-surround feature, we make a assumption that if the area ratio of the minimum rectangle of region cluster is larger than 0.7, the region is regard as background. But sometimes the area ratio of minimum rectangle of salient object cluster is more than 0.7, so it will mistake salient objects as background. For the color-distribution feature, sometimes area ratio or position of the salient object is not in the hotspot of histogram, it may not be detected as a salient object. Some results are shown in Fig. 18, (a) is original image, (b) is center-surround salient map, (c) is color-distribution salient map and (d) is the combination salient map. In practice, only use one of the features may cause unsatisfactory results. In row (1), the

Fig. 15 Center-Surround salient map

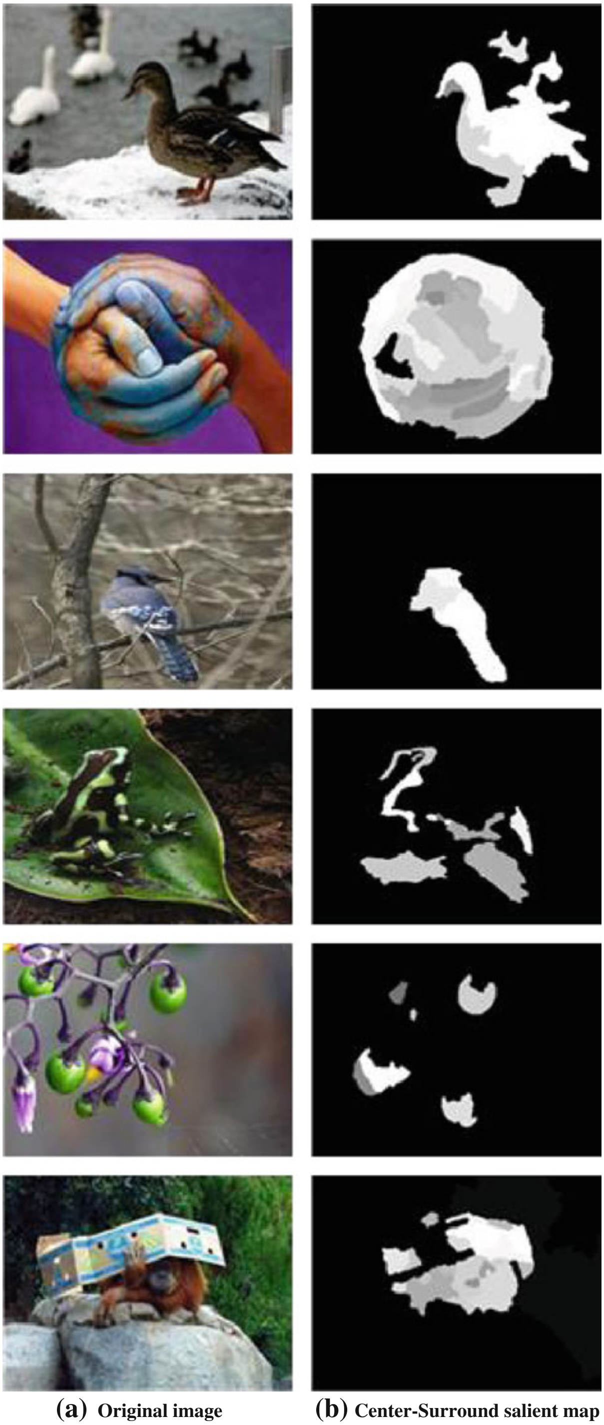
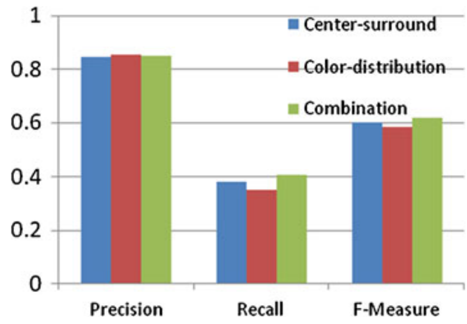


Fig. 16 Center-Surround salient map

(a) **Originalimage** **(b)** **Color-Distribution saliencyy map**

Fig. 17 Quantitative results of using different saliency features



center-surround salient map obviously localizes the incomplete visual attention, but we obtain the better color-distribution salient maps. In row (2), in contrast to the first example, the saliency map of color-distribution feature works not well but the saliency map of center-surround feature looks good. Sometimes any of them can only get part of the salient object; we must combine the two salient maps to get a complete salient map, as in row (3, 4). The results of the experiment show that the combination of them can achieve better performance, the quantities results are shown in Fig. 17.

Generally, we use grayscale to process an image. For multi-color images, we can combine the salient maps of different color channels to get multi-color salient objects.

4.4 Experiment 2: Comparison with color-orientation framework

Our approach and Color-Orient Framework [33] use the same MSRA image database to evaluate posed approaches. The saliency map comparison of the proposed method and Color-Orientation Framework on selected images from MSRA image database is shown in Fig. 19. It is evident that the proposed method outperforms that method. For example, in row (2), our method detects the windmill and one flabellum as saliency object, but [33] can only detect one flabellum. We don't use low level feature such as pixel etc. We use region as basic compute unit, so the salient map we get is continuous and integrated, such as in row (3, 4, 5, and 6). We use area and position global features, so we can get salient map more accurate. Take row (7) for example, the proposed method can detect the crocodile, but [33] fails.

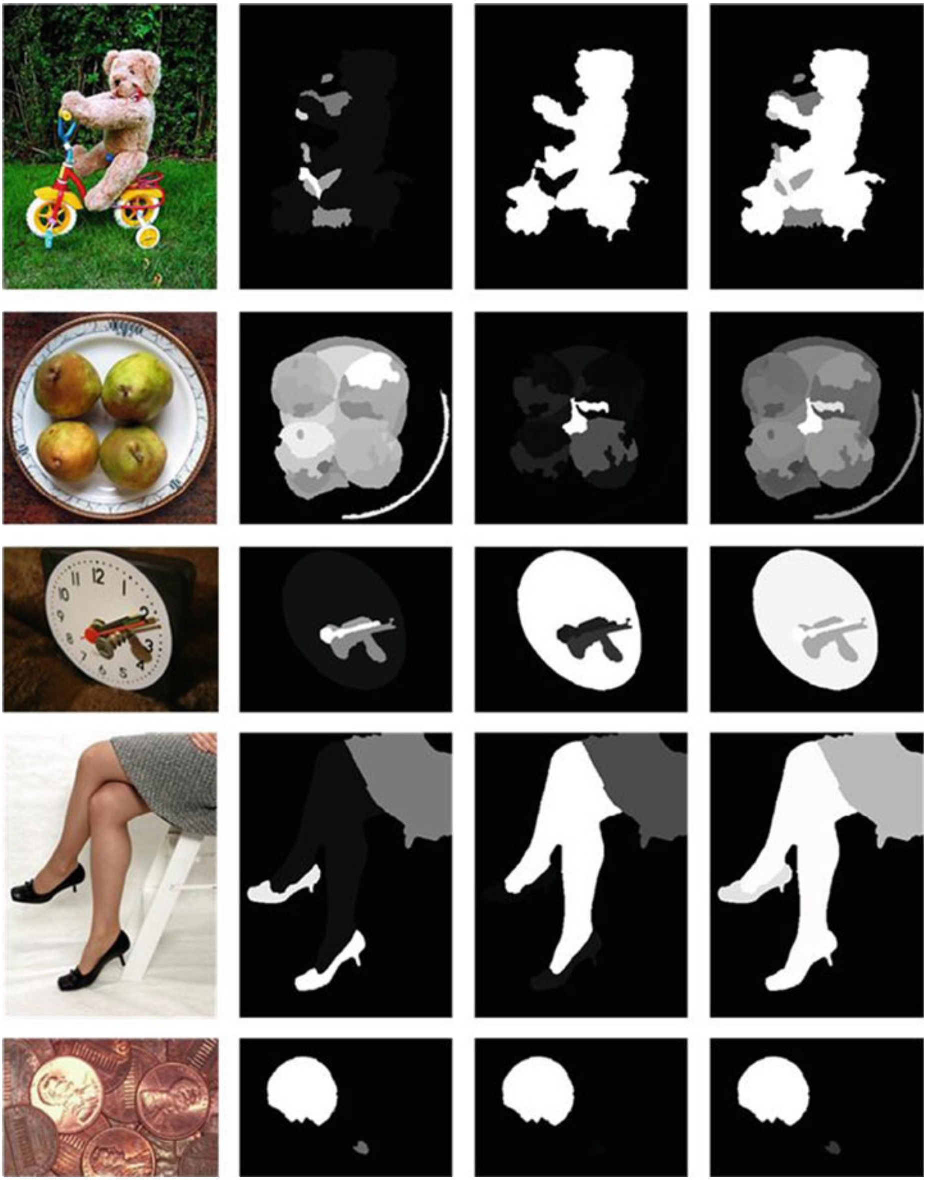
4.5 Experiment 3: Comparison with other approaches

We also compare our approach with four existing approaches: Saliency Tool Box [34], Contrast Method [18], Spectral Residual Method [4], and Color-Orientation Framework [33]. In the evaluation part of [33], the experimental results of four approaches are listed. In this paper, we use the same image dataset and evaluation method as [33], and compare with the results in [33] directly. From Fig. 20, we can see our approach has better performance, especially *Precision*.

4.6 Experiment 4: Analysis of robustness

We discuss the robust of our approach.

Our approach is robust against background noise. The result is not influenced by different parameters. In region clustering, we use clustering threshold. If the distance of two regions is



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

Fig. 18 Results of salient object detection by combining two types of saliency maps **a** Original image. **b** Center-Surround saliency map. **c** Color-distribution saliency map. **d** Combination of (**b**) and (**c**)

less than threshold, we cluster the two regions. We randomly choose 15 for the threshold value. But it doesn't influence the result. We can give it from 12 to 18; the result will not be influenced by it. It is shown as Table 2.

We train 1000 images to get the probability histogram of area and position and test other 4000 images in our approach. From Fig. 3 Probability histogram of positions and Fig. 6

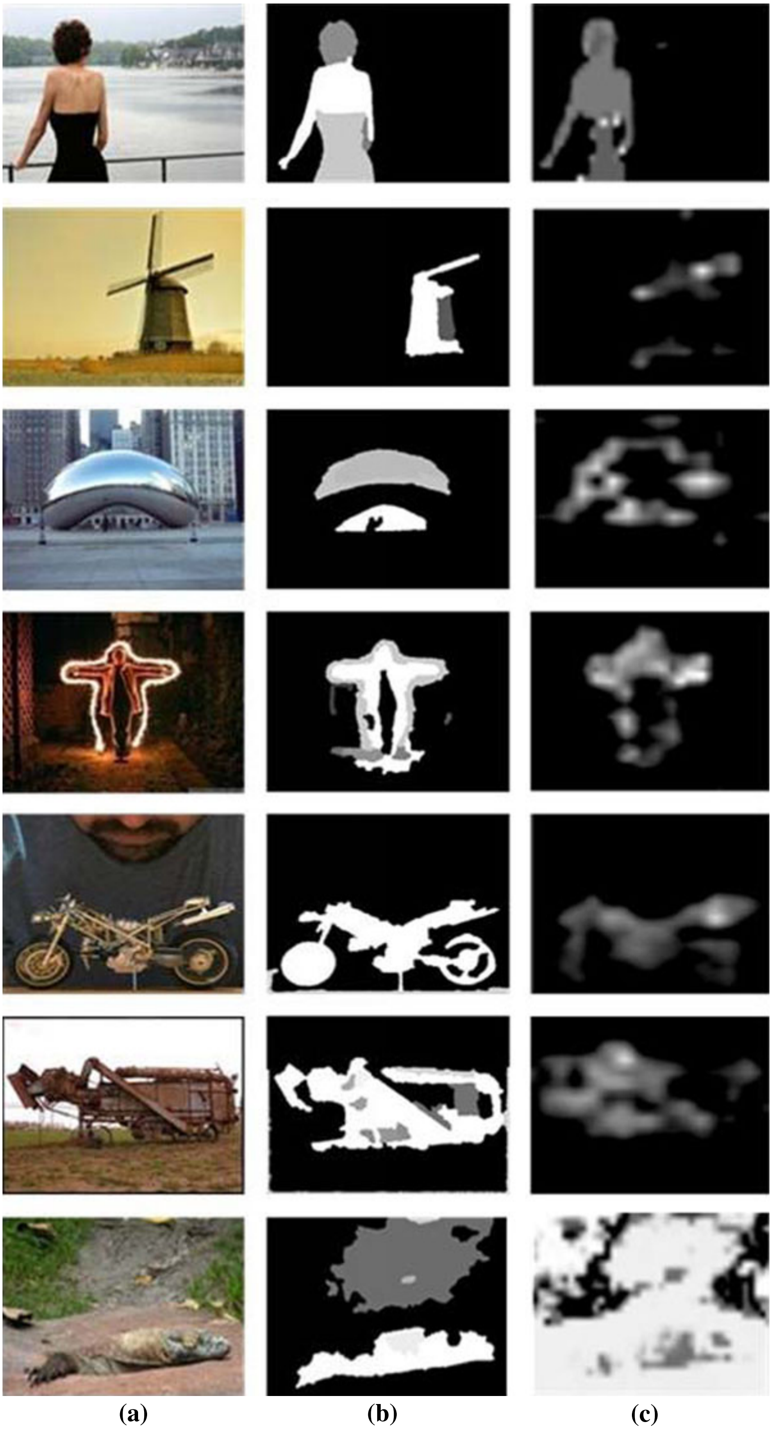
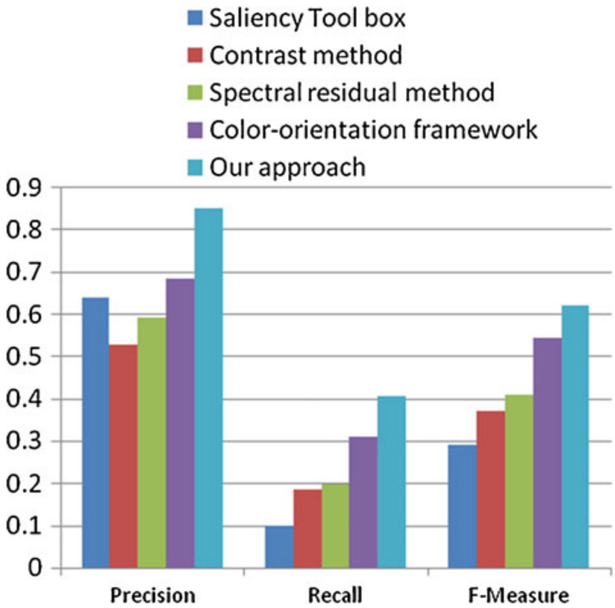


Fig. 19 Comparison with Color-Orientation Framework (**a**) Original images (**b**) Our method (**c**) Color-Orientation Framework

Fig. 20 Comparison of average precision, recall, and F-measure



Probability histogram of area ratio, we can see that they follow fixed law. We can choose other 1000 images to train the data, the result is consistent. The result is shown in Table 3.

4.7 Failure cases and analysis

Although the new method performs well in many cases, it may fail under some conditions. For example in the image of the first row of Fig. 21, the grassland finally becomes the salient object because of its high color contrast with its surrounding areas, as well as its moderate size and central location. In the second row of Fig. 21, the human body region cannot be detected correctly because the color of clothes is almost the same as the background. These failures arise from that our method does not cope with the high-level semantic features of objects. In some conditions, the analysis solely based on color, area and position information is not accurate enough to find out the salient object in the image. We will focus on solving these problems in our future work.

Table 2 Comparison of different Cluster Thresholds

Cluster Threshold	Precision	Recall	F-Measure
12	0.833	0.430	0.626
13	0.842	0.428	0.630
14	0.846	0.424	0.627
15	0.851	0.412	0.621
16	0.858	0.397	0.617
17	0.858	0.393	0.613
18	0.865	0.409	0.626

Table 3 Comparison of different Train Images

Train Images	Precision	Recall	F-Measure
0001-1000	0.851	0.412	0.621
1001-2000	0.838	0.420	0.624
2001-3000	0.853	0.400	0.624
3001-4000	0.849	0.399	0.610
4001-5000	0.842	0.387	0.604

4.8 Efficiency analysis

With regard to efficiency, the region-based computation takes much less time than pixel-based approaches because our approach is based on regions and the number of regions is generally only a few dozen blocks,. The time cost concentrates on the stage of image segmentation. The time complexity of many existing algorithms for image segmentation is about $O(n^2)$, and some algorithms can achieve $O(n \times \log(n))$ [25], where n is the total number of pixels in the image. The time complexity of our algorithm is very close to the recently proposed algorithms for salient object detection, and better than some pixel-based approaches. The conclusion can be drawn that our approach improves performance of salient object detection while ensuring the efficiency of detection. Our approach can provide similar quality at higher performance, or better quality for the same performance.

5 Conclusion and future work

The salient object detection method proposed in this paper is based on regions. According to the human visual perception, pixels in homogeneous regions obtained by image segmentation should have the same saliency. So regions are used as basic units to compute the saliency in an image. The salient object is a combination of homogeneous regions, and the final output can be obtained by combining the homogeneous regions.

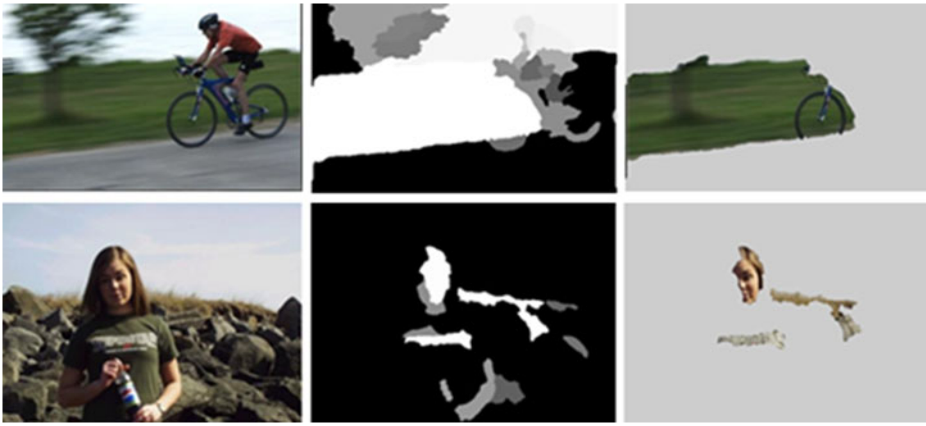


Fig. 21 Some failures of detection

Through a visual saliency survey and experiments on a large image set, we conclude how area and position have impacts on saliency. Then we propose the region-based center-surround feature and region-based color-distribution feature as new saliency features based on size and location. The region-based center-surround feature can accurately determine the shape and position of the central region thus reflect the difference of the target and its surroundings more accurately. The region-based color-distribution feature applies area clustering to analyze the distribution of homogeneous regions, then computes saliency by area and position saliency features. The saliency of homogeneous regions reflects the distribution of color in the image.

References

1. Cheng H-D, Sun Y (2000) A hierarchical approach to color image segmentation using homogeneity. IEEE Trans Image Process 9(12):2071–2082
2. Deng Y, Manjunath B (2001) Unsupervised segmentation of color-texture regions in images and video. IEEE Trans Pattern Anal Mach Intell 23(8):800–810
3. Frintrop S, Backer G, Rome E (2005) Goal-directed search with a top-down modulated computational attention system. DAGM-Symposium:117–124
4. Hou X, Zhang L (2007) Saliency Detection: A Spectral Residual Approach. IEEE Conference on Computer Vision and Pattern Recognition (CVPR):1–8
5. Itti L, Koch C (2001) Computational modeling of visual attention. Nat Rev Neurosci 2(3):194–203
6. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20(11):1254–1259
7. Kanan C, Tong M, Zhang L, Cottrell G (2009) SUN: Top-down saliency using natural statistics. Visual Cognition 17:979–1003
8. Kim S, Park S, Kim M (2003) Central object extraction for object-based image retrieval. Lecture Notes in Computer Science on Image and Video Retrieval l(2728):523–528
9. Ko BC, Byun HR (2003) Region-based image retrieval: a new method for extraction of salient regions and learning of importance scores. International Journal of Pattern Recognition and Artificial Intelligence 17(8):1349–1367
10. Koch C, Ullman S (1985) Shifts in selective visual attention: towards the underlying neural circuitry. Hum Neurobiol 4:219–227
11. Liansheng Z, Ketan T, Nenghai Y, Yangchun Q (2009) Fast salient object detection based on segments. International Conference on Measuring Technology and Mechatronics Automation 1:469–472
12. Lin L, Liu X, Zhu S-C (2010) Layered graph matching with composite cluster sampling. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 32(8):1426–1442
13. Lin L, Luo P, Chen X, Zeng K (2012) Representing and recognizing objects with massive local image patches. Pattern Recognition 45(1):231–240
14. Lin L, Peng S, Porway J, Zhu S-C, Wang Y (2007) An empirical study of object category recognition: sequential testing with generalized samples. Proc IEEE International Conference on Computer Vision (ICCV) 1:419–426
15. Lin L, Wu T, Xu Z, Porway J (2009) A stochastic graph grammar for compositional object representation and recognition. Pattern Recognition 42(7):1297–1307
16. Liu T, Sun J, Zheng NN, Tang X, Shum H-Y (2007) Learning to detect a salient object. IEEE Conference on Computer Vision and Pattern Recognition (CVPR):1–8
17. Lucchese L, Mitra SK (2001) Color image segmentation: a state-of-the-art survey. The Indian National Science Academy (INSA-A) 67(2):207–221
18. Ma Y-F, Zhang H-J (2003) Contrast-based Image Attention Analysis by Using Fuzzy Growing. ACM International Conference on Multimedia:374–381
19. Marchesotti L, Cifarelli C, Csurka G (2009) A framework for visual saliency detection with applications to image thumb nailing. International Conference on Computer Vision (ICCV):2232–2239
20. Meur O, Callet P, Barba D, Thoreau D (2006) A coherent computational approach to model bottom-up visual attention. IEEE Trans Pattern Anal Mach Intell 28(5):802–817
21. Navalpakkam V, Itti L (2006) An integrated model of top-down and bottom-up attention for optimizing detection speed. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2:2049–2056
22. Nothdurft H (2000) Salience from feature contrast: add itivity across dimensions. Vision Res 40 (10):1183–1201

23. Nothdurft H (2000) Salience from feature contrast: variations with texture density. Vision Res 40 (23):3181–3200
24. Ouerhani N, Bracamonte J, Hugli H, Ansorge M, Pellandini F (2001) Adaptive color image compression based on visual attention. International Conference on Image Analysis and Processing (ICIAP):416–421
25. Pedro F, Daniel P (2004) Efficient graph-based image segmentation. International Journal of Computer Vision 59(2):167–181
26. Shi J, Malik J (2000) Normalized cuts and image segmentation. IEEE Trans Pattern Analysis and Machine Intelligence 22(8):888–905
27. Stentiford F (2007) Attention based Auto Image Cropping. Workshop on Computational Attention and Applications (ICVS)
28. Sun Y, Fisher R (2003) Object-based visual attention for computer vision. Artificial Intelligence 146 (1):77–123
29. Thoma G, Hichem S (2007) Region-oriented visual attention framework for activity detection. The 4th International Workshop on Attention in Cognitive Systems (WAPCV): 481–496
30. Thomas G, Hichem S, David H, Iris V, Valentin E (2008) Modeling attention and perceptual grouping to salient objects. The 5th International Workshop on Attention in Cognitive Systems (WAPCV):166–182
31. Treue S (2003) Visual attention: the where, what, how and why of saliency. Curr Opin Neurobiol 13 (4):428–432
32. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol: 511–518
33. Viswanath G, Yiqun H, Deepu R (2009) Salient region detection by modeling distributions of color and orientation. IEEE Transactions on Multimedia 11(5):892–905
34. Walther D, Koch C (2006) Modeling attention to salient proto-objects. Neural Network 19(9):1395–1407
35. Zhang W, Wu Q, Wang G (2008) Adaptive semantic bayesian framework for image attention pattern recognition. Proceedings of the International Conference on Pattern Recognition (ICPR):1–4

Zhuojia Liang received the B.Sc. degree from the School of Computer Science and Engineering, South China University of Technology, China in 2008, and received the M.Sc. degree from the School of Information Science and Technology, Sun Yat-Sen University, China in 2010. His research interests include image salient object detection and CAPTCHA.

Mingjia Wang received the B.Sc. degree from the School of Information Science and Technology, Sun Yat-Sen University, China in 2009. He now is a postgraduate student at the School of Information Science and Technology, Sun Yat-Sen University, China. His research interests include image inpainting and image salient object detection.

Xiaocong Zhou received the BS and MS degrees from Sun Yat-Sen University in 1992 and 1995, respectively. He received the PhD degree from Institute of Software, Chinese Academy of Sciences in 2001. He is now an associate professor, Department of Computer Science, Sun Yat-Sen University. His research interests include software testing and debugging, and the theory and technology of software engineering.

Liang Lin received his B.S. in Electronics Science and Technology, Beijing Institute of Technology, Beijing, China in 2003. He received the Ph.D in Computer Science and Technology, Beijing Institute of Technology, Beijing, China in 2008. He is now an associate professor of Computer Science, Sun Yat-Sen University. His research interests include computer vision and graphics and image/video processing.

Wenjun Li received the BS and MS degrees from Sun Yat-Sen University in 1988 and 1991, respectively. He received the PhD degree from the Institute of Software, Chinese Academy of Sciences in 2002. He is currently a professor of Computer Science and Software Engineering with Sun Yat-Sen University. He is the dean of the School of Software, Sun Yat-Sen University. His research interests include but are not limited to software testing and debugging, service computing, models for concurrency, and computer-human interaction.