SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

March 2016, Vol. 59 032104:1–032104:14
doi: 10.1007/s11432-015-5420-9

Salient object detection via region contrast and graph regularization

Xingming WU, Mengnan DU, Weihai CHEN* & Jianhua WANG

School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191*, China*

Received April 1, 2015; accepted June 24, 2015; published online January 18, 2016

Abstract Detection of salient objects in an image is now gaining increasing research interest in computer vision community. In this study, a novel region-contrast based saliency detection solution involving three phases is proposed. First, a color-based super-pixels segmentation approach is used to decompose the image into regions. Second, three high-level saliency measures which could effectively characterize the salient regions are evaluated and integrated in an effective manner to produce the initial saliency map. Finally, we construct a pairwise graphical model to encourage that adjacent image regions with similar features take continuous saliency values, thus producing the more perceptually consistent saliency map. We extensively evaluate the proposed method on three public benchmark datasets, and show it can produce promising results when compared to 14 state-of-the-art salient object detection approaches.

Keywords salient object detection, region contrast, region compactness, global distinctness, graph regularization

Citation Wu X M, Du M N, Chen W H, et al. Salient object detection via region contrast and graph regularization. Sci China Inf Sci, 2016, 59(3): 032104, doi: 10.1007/s11432-015-5420-9

1 Introduction

Visual attention is a particularly important aspect of human visual system. It enables a person to perceive salient regions in complex scenes quickly and then selectively evaluate the small pieces of information. This ability to extract the most relevant (salient) sensory information during an early processing stage considerably enables effective understanding and rapid reaction in a complex world. Visual saliency has attracted considerable attention during the last few decades, and has been studied extensively by researchers in the fields of neurobiology [1], cognitive psychology [2] and computer vision [3–7]. Salient object detection has a broad range of applications in the fields of computer vision and graphics, including object detection and segmentation [8], image retrieval [9], image registration [10], image cropping [11], and so on.

Early computational visual saliency studies focus on fixation prediction. The goal of fixation prediction is to evaluate a probabilistic map of an image that simulates human eye movement activities. Recently, visual saliency has been extended to salient object detection for its potential application in other machine vision fields. This type of model aims to detect regions in a scene that may contain salient objects. This

* Corresponding author (email: whchenbuaa@126.com)

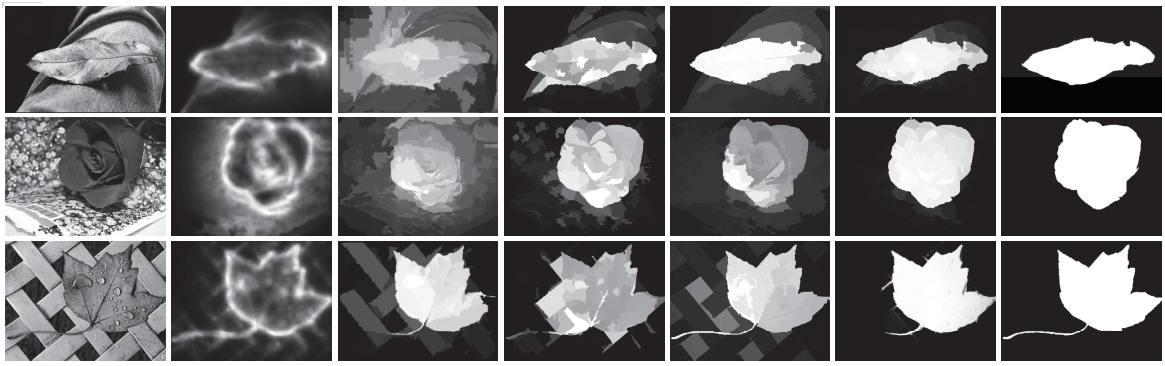


Figure 1 Saliency detection results using state-of-the-art methods as well as ours. (a) Input images; (b) [12] merely captures object boundary (CA); (c) [4] produces saliency map with ill-defined boundary (RC); (d) [13] some background regions are incorrectly detected as salient (GS); (e) [14] detects parts of background region as salient (HS); (f) our results uniformly highlight saliency region and effectively exclude background pixels; (g) ground truth masks (GT).

is also the focus of this study. In generally, a "saliency map" is generated where the intensity of its pixel represents the possibility of that pixel belonging to the salient object. The entire object is then segmented from the saliency map using various methods (e.g., simple thresholding).

Saliency detection methods generally follow the center-surround contrast principle of [15]. Although much progress has occurred regarding the performance of these methods, it remains a problem to detect salient regions from the complex images. These high-contrast textures and confusing patterns in the background may be inaccurately detected as salient by these contrast based methods. Figure 1 presents three scenes of increasing background complexity. The detection results yielded from state-of-the-art methods either outline object boundaries but fail to extract the interior as shown in Figure 1(b), or have ill-defined object boundaries as shown in Figure 1(c). The fuzzy results limit the usefulness of these methods in many applications, such as image segmentation and object recognition.

These results indicate that merely applying the contrast principle is not sufficient for saliency detection. Thus, some researchers [5, 13] attempt to adopt some prior knowledges to address this problem. In [13], boundary and connectivity prior were used as additional cues of saliency detection. However, this type of method may not always function well, because these priors may be valid when the objects are placed near the image boundaries. For example, some background regions are incorrectly detected as salient by [13] in Figure 1(d). Based on our analysis, we determine that an effective saliency detection method should consider the following: (1) Image decomposition. The input image is decomposed into compact and regular regions. This process can remove image redundancies and reduce the complexity of subsequent saliency cues evaluation. (2) Region contrast. More discriminative features may be extract from the regions than pixels. Besides, the small number of regions also ensures efficiency. (3) High-level priors. These can effectively characterize salient and background regions, yielding accurate saliency detection. (4) Spatial context optimization. A pairwise graphical model is constructed to smooth raw saliency maps and augment the dissimilarity between the object and background regions.

Following these four principles, we propose a novel salient object detection method. The proposed approach first decomposes the image into regular-size regions that meet object boundaries. We then calculate the compactness of a region using its spatial distribution in terms of color, in order to seek regions that have high possibility of containing a meaningful object. After segmenting the region compactness map, we obtain the soft foreground regions. A novel global distinctness saliency map is generated by making comparisons with the soft foreground regions. The spatial location prior is then used to increase the difference between the foreground and background. These three cues, i.e., region compactness, objectness-guided global distinctness and spatial location prior, are integrated to produce the initial saliency map. Finally, we use a graph-based regularization method that is modified from a pairwise CRF framework to refine the initial saliency values. The purpose is to produce more consistent saliency maps. We evaluate

the performance of the proposed method extensively on three well-known datasets and our approach produces promising results compared to those of state-of-the-art salient object detection algorithms.

The main contribution of this study is two-fold. The first contribution involves three high-level saliency measures that may effectively characterize a salient object and dramatically increase the accuracy and robustness of saliency detection methods. The second contribution is an efficient graph regularization measure, which models the spatial context relationship of the initial saliency map and eventually enhances the coherence between the salient regions.

The rest of the paper is organized as follows. Section 2 gives a brief review of related saliency detection methods. Section 3 presents the proposed algorithm which generates the initial saliency map. Section 4 introduces a graph based optimization measure to refine the initial saliency map. Experimental results and comparisons are described in Section 5. Conclusion and future work are presented in Section 6.

2 Related work

Visual saliency models may be categorized into three types: fixation prediction, salient object detection and objectness. Fixation prediction models aim to predict where human look in images. Salient object detection methods are designed to extract the most attention-grabbing objects in a scene. Objectness algorithms act as a class-generic object detector, which quantifies the possibility of an image window containing an object of any class [16]. In the following, we briefly review the type of salient object detection that are most relevant to our approach.

Salient object detection models can be classified into pixel-based methods and region-based methods, depending on whether the contrast evaluation is defined over pixels or regions. Pixel-based methods calculate center-surround color contrast of a pixel with its local neighborhood or the whole image pixels via some features (e.g., color, intensity and orientation). Achanta et al. [17] determined salient region employing local contrast at various scales using low-level color and luminance features. Goferman et al. [12] incorporated local low-level cues, global information, center priors and high-level factors to compute a context aware saliency map. Achanta et al. [3] proposed a frequency-tuned method by measuring the differences between the feature of individual pixel and the average feature of the entire image. However, these methods tend to produce saliency maps that highlight the edges rather than the entire salient object, making them less useful in many applications.

Region-based approaches evaluate color uniqueness over image regions instead of pixels. Cheng et al. [4] extended the saliency estimation from pixel-wise to patch-level by calculate the contrast of a segmentation patch with its surrounding patches. More recently, Perazzi et al. [18] formulated the region contrast evaluation into Gaussian filters, remarkably reducing the running time of saliency detection. Yan et al. [14] analyzed saliency cues from multi-scale structures in order to tackle small-scale patterns detection problem. Region-based methods possess two advantages over pixel-based ones. On one hand, methods could benefit from the high-level features extracted from regions, and could highlight the whole extent of object. On the other hand, the relatively small number of regions could boost the efficiency of saliency detection algorithms, accelerating the practical application process of saliency approaches in other fields.

Contrast-based saliency detection methods usually have poor performance in complex scenes. Thus, lots of domain knowledge from other fields have been applied to the saliency estimation. In [19], Han et al. explored the properties of the foreground regions and proposed a probabilistic computational model by integrating objectness likelihood with appearance rarity for visual saliency detection. The background priors that treates the image boundaries as pseudo background were studied in [5, 13, 20]. Han et al. [21] learned latent patterns from the background using the deep learning architecture to separate salient regions from complicated images. The influence of heterogeneous background on visual saliency was studied in [22]. Wu and Shen [23] utilized low-rank and sparse matrix decomposition methods for saliency detection. Liu et al. [24] learned a Partial Differential Equations (PDE) system from an image for binary saliency estimation. Chang et al. [25] constructed a graphical model to integrate objectness and saliency.

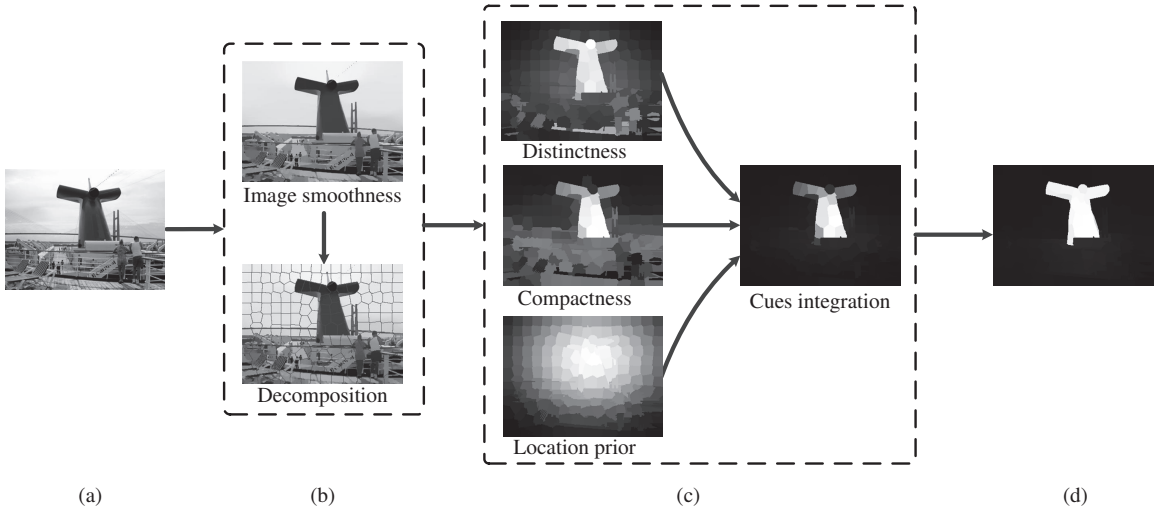


Figure 2 The architecture of our pipeline. (a) The input image; (b) the image decomposition; (c) the calculation of three saliency measures and the saliency integration; (d) the final saliency map after saliency optimization.

Li et al. [26] explored the combination of fixation prediction and salient object segmentation.

There also exist some learning-based methods in the literature. The task of salient region detection are generally formulated as a binary labeling problem. Khuwuthyakorn et al. [27] combined salient features using a mixture of support vector machines (SVMs) to extract salient object from an image. Liu et al. [28] extracted salient regions from both a picture and sequential images by using CRF learning. Jiang et al. [6] modeled saliency estimation as a regression problem and learned to integrate different regional features employing a random forest. Although impressive performances have been realized by these methods, they usually needs time-consuming training, which is not feasible in many applications.

3 Proposed approach

The proposed method consists of three procedures. First, the input image is segmented into super-pixel regions. Second, we use three saliency estimation measures to calculate the saliency value of each region. These saliency measures include region compactness, objectness-guided global distinctness and spatial location prior. A two-layer framework is then employed to integrate these three measures into a master saliency map. Finally, we construct a graphical model to refine the integrated saliency values and produce the more consistent saliency map. Figure 2 shows the pipeline of the proposed method. Details of these steps are provided in the following sections.

3.1 Image decomposition

The edge-preserving smoothness measure in [29] is first used to smooth the input image. It effectively removes the small-amplitude details and retains the high-contrast foreground object edges in the meantime. To capture further the structural characteristics of the smoothed image, a color-based super-pixels segmentation approach [30] is used to decompose the image into regions that have regular size and meets the object boundaries well. This structural model $\mathcal{V}$:

$$\mathcal{V} = \{\nu_1, \nu_2, \nu_3, \ldots, \nu_K\}, \tag{1}$$

which contains K image regions effectively alleviates the influence of cluttered noises in images. Besides, this image decomposing measure guarantees the computation efficiency of subsequent saliency measures.

Unlike previous decomposition techniques that rely on large-scale segmentation [4, 14], our approach decomposes images into super-pixels. This kind of representation is less likely to cross object boundaries,

(a) (b)

Figure 3 Region compactness measure: (a) input image; (b) region compactness saliency map. We notice that compactness measure could effectively characterize salient object and clearly distinguish it from the background.

thus generating high quality segmentations. Moreover, using the regular super-pixel as the unit for saliency estimation facilitates the integration of different saliency cues in subsequent calculation procedures. As shown in Figure 2(b), our soft abstraction approach clusters pixels into compact and boundary-preserving super-pixel regions. We fix the number of super-pixel regions to 200 to balance accuracy and efficiency.

3.2 Region compactness

Low-level saliency cues (e.g., color contrast) are generally known to perform poorly in data-driven salient object detection. Therefore, introducing some high-level saliency priors in order to characterize salient regions effectively is essential. We observe that the colors belonging to the background generally spread over the entire image, while colors belonging to the foreground object concentrate within a small part of regions. That is to say, there is a larger probability of being salient for regions with compact color distribution or less color variance in the spatial domain [31]. Therefore, we adopt region compactness as our first saliency measure. The compactness score C_k of region k is defined as the inverse of its color spatial variance:

$$C_k = (\sum_{j=1}^{K} w_{\text{color}}(c_k, c_j) \cdot d^2_{\text{spa}}(p_j, u_k))^{-1}, \quad k \in \mathcal{V}. \tag{2}$$

The Gaussian function is used as weight to measure the color similarity of two regions:

$$w_{\text{color}}(c_k, c_j) = \frac{1}{z_k} \cdot \exp(-\frac{d^2_{\text{color}}(c_k, c_j)}{2\mu_c^2}), \tag{3}$$

where z_k is the normalization term that guarantees $\sum_{j=1}^{K} w_{\text{color}}(c_k, c_j) = 1$. $d_{\text{color}}(c_k, c_j)$ measures the color distance of two regions in the Euclidean space. Parameter μ_c^2 controls the degree of color similarity, it is set to 20 in all experiments. The $d_{\text{spa}}(p_j, u_k)$ in (2) denotes the spatial distance between the position centroid of two regions in the Euclidean space. u_k in (2) is the weighted mean position of region k:

$$u_k = \sum_{j=1}^{K} w_{\text{color}}(c_k, c_j) \cdot p_j. \tag{4}$$

We present an example in Figure 3 to demonstrate the effectiveness of our compactness measure. In Figure 3(a), the two purple petunias are more compact compared to the brown mood beneath them, thus should be considered more salient, as demonstrated from the compactness saliency map in Figure 3(b). It is proven to be an effective measure to capture object information and accurately distinguishes salient regions from background. Besides, our compactness measure can uniformly highlight salient object for taking into consideration the relative spatial information.

3.3 Objectness-guided global distinctness

The distinctness of an image color is considered the most important contributor to visual saliency [31,32]. Foreground regions usually possess more distinct colors than the background regions. Global regional

contrast is generally measured by calculating the color differences of different image regions in a global manner. While the global contrast method may detect regions with distinct color, it can't handle all kinds of images, especially when the foreground objects possess similar color to the background regions. If the rough position of the salient regions could be located before calculating the global regional contrast, the robustness of this saliency cue would increase. We segment the region compactness saliency region using an adaptive threshold to obtain the "soft" foreground and background. The threshold is set to twice the mean region compactness value of an image. The objectness likelihood of super-pixel region $j \in \mathcal{V}$ is defined as

$$o(\nu_j) = \begin{cases} 1, & \text{if } C_j > \sum_{i=1}^{K} \frac{2C_i}{K}, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

As the segmentation is coarse, we measure the contrast between regions $k \in \mathcal{V}$ and the "soft" foreground regions, aiming to suppress the regions inside the "soft" foreground similar to the outside part, while highlighting the outside regions similar to the inside part. Based on this principle, we define the objectness-guided global distinctness as our second saliency measure. Given K regions, the objectness-guided global distinctness D_k of region k is calculated as

$$D_k = \sum_{j=1}^{K} w_{\text{spa}}(p_k, p_j) \cdot \frac{o(\nu_j)}{d^2_{\text{color}}(c_j, c_k)}, \quad k \in \mathcal{V}. \tag{6}$$

We use Gaussian weight $w_{\text{spa}}(p_k, p_j) = \exp(-\frac{d^2_{\text{spa}}(p_k, p_j)}{2\mu_d^2})$ to measure the spatial distance of region k and region j. Parameter μ_d^2 controls the range of distance, and we empirically fix it to 0.2.

The proposed objectness-guided global distinctness measure assigns larger saliency value to regions whose color is less distinct to the "soft" foreground regions or whose spatial distance is closer to the "soft" foreground regions. Note that we calculate the objectness-guided global distinctness procedure in CIELab color space, because it is perceptually accurate. In addition, this measure is calculated in region-level, which guarantees the efficiency of this procedure.

3.4 Spatial location prior

Photographers tend to align object of interest at the center of photos. Thus, we employ this prior to depress the saliency of background pixels and define it as our third saliency measure. Some literatures have studied the effect of center prior on saliency estimation [33, 34]. However, previous center priors assign larger saliency weight to the centroid of an image [12]. It may fail to function when salient objects are away from the image center. We present a more accurate spatial location prior to solve this problem. Because the proposed region compactness maps give the rough location of the salient region, the compactness saliency score is used as the weight to estimate the object center p_{center}. The definition is

$$p_{\text{center}} = \frac{1}{M} \sum_{j=1}^{K} D_j \cdot p_j, \tag{7}$$

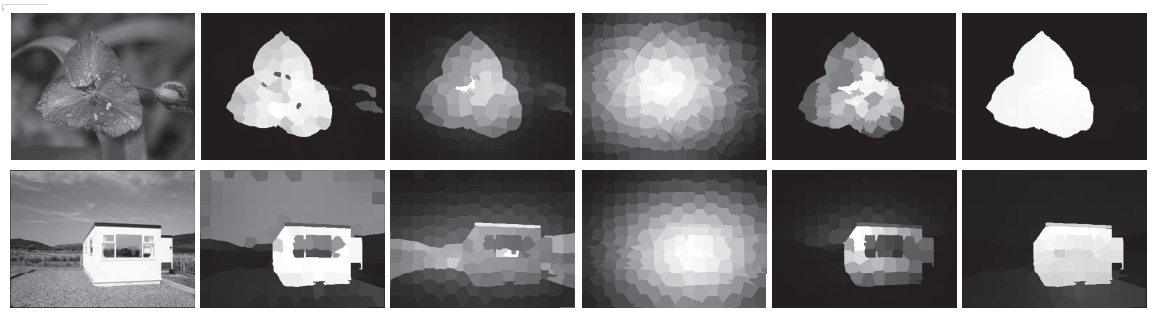
where $M = \sum_j D_j$ is the normalization term that ensures all the weights add up to one, D_j is the region compactness saliency score of j, P_j is the mean position of region j. Our spatial location prior L_k of region k is evaluated by measuring the distance of the region to the estimated object center P_{center}:

$$L_k = \exp\left(- \frac{d^2_{\text{spa}}(p_k, p_{\text{center}})}{2\mu_m^2} \right), \quad k \in \mathcal{V}, \tag{8}$$

where parameter μ_m^2 represents the range of distance, we fix μ_m^2 to 0.15 in all experiments.

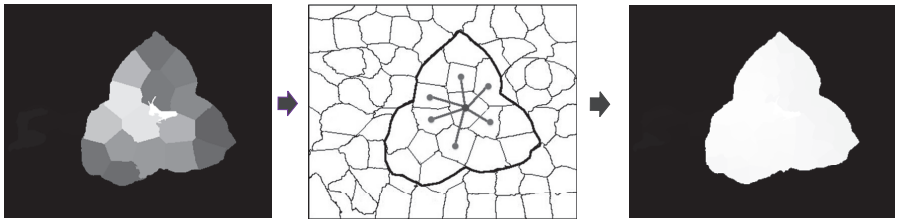
3.5 Integration

The aforementioned three saliency measures perform differently in saliency detection, as shown in Figure 4 (b)–(d). Therefore, it is essential to integrate these measures to obtain an accurate master saliency map.



(a) (b) (c) (d) (e) (f)

Figure 4 Saliency map of different procedure. (a) The input image; (b) region compactness map; (c) objectness-guided global distinctness map; (d) spatial location prior map; (e) the two-layer integration saliency map; (f) the final saliency map employing a graphical model.



(a) (b) (c)

Figure 5 Spatial context optimization via graph regularization. We adopt the graph regularization measure which is modified from a pairwise CRF framework (b) to model spatial context relationships of the initial saliency map (a), and get the more consistent saliency map (c).

Motivated by the saliency combination framework in [35], we introduce a similar two-layer integration structure. It is made up of two layers: the base layer and the enhancement layer. They are elaborately described in the following.

• Base layer: In this layer, we seek regions both salient in compactness and distinctness. To this end, we investigate two approaches to integrate these two cues. The first approach is weighted sum of individual cues, i.e., $B_k = w_1 C_k + w_2 D_k$. The second integration scheme uses multiplication, i.e., $B_k = C_k \cdot D_k$. We empirically select the latter, for it could better assign higher salient value to salient regions and depress the background regions.

• Enhancement layer: It corresponds to spatial location prior measure. This layer complements the base layer by assigning larger saliency value to object center. As such, the proposed approach further enlarges the dissimilarity between foreground and background regions.

We first normalize the region compactness C_k, objectness-guided global distinctness D_k and spatial location prior L_k to the range [0,1], then we use the proposed structure to integrate the three measures. The master saliency score $\tilde{S}_k$ of region k is calculated as the product of the base layer and the enhancement layer:

$$\tilde{S}_k = B_k \cdot L_k = C_k \cdot D_k \cdot L_k, \quad k \in \mathcal{V}. \tag{9}$$

After the saliency integration of the complementary two layers, we get a more accurate saliency map (see Figure 4(e)).

4 Spatial context optimization via graph regularization

The saliency integration map $\tilde{S}$ is limited in that saliency values are not continuous even between adjacent regions with similar patterns (see Figure 5(a)). We intend to produce the consistent saliency map S (see Figure 5(c)) where the entire salient regions are uniformly highlighted, object boundary discontinuity is well preserved and background pixels are effectively suppressed. From this point of view, saliency

estimation can be modeled as a foreground labeling task, i.e., assigning value closer to 1 to object region and value closer to 0 to background region. Towards this goal, we adopt a modified Conditional Random Fields (CRF) framework [36] to model spatial context relationships and enforce the consistence of the initial saliency map.

In the CRF framework, given the observation *priori* D, the posterior distribution S is formulated as *Gibbs* distribution:

$$P(S|D) = \frac{1}{Z} \cdot \exp(-E(S|D)), \tag{10}$$

where Z is normalization constant which is called the *partition function*. The corresponding Gibbs energy is written as $E(S) = -\log P(S|D) - \log Z$. Generally, an undirected graph $G = (\mathcal{V}, \mathcal{N})$ is constructed, where $\mathcal{V}$ is the set of random variables. The neighborhood system $\mathcal{N}$ of random field is defined by the sets $\mathcal{N}_i$, $\forall i \in \mathcal{V}$, where $\mathcal{N}_i$ denotes all neighbors of the variable S_i. In this study, we use a pairwise CRF whose energy can be written as the sum of unary and pairwise potentials as

$$E(S) = \sum_{i \in \mathcal{V}} \psi_i(S_i) + \sum_{i \in \mathcal{V}, j \in \mathcal{N}_i} \phi_{ij}(S_i, S_j), \tag{11}$$

The Maximum a Posteriori(MAP) labeling S^* of random field is

$$S^* = \arg\max P(S|D) = \arg\min E(S). \tag{12}$$

In our modified framework, the set of random variables $S = \{S_1, S_2, \ldots, S_K\}$ correspond to super-pixels $\mathcal{V} = \{\nu_1, \nu_2, \nu_3, \ldots, \nu_K\}$. The neighborhood $\mathcal{N}_i$ of random variable $\forall i \in \mathcal{V}$ denotes all adjacent super-pixels of the variable S_i (see Figure 5(b)). The image color sets $C = \{C_1, C_2, \ldots, C_K\}$ represent the observation priori $D = \{D_1, D_2, \ldots, D_K\}$. The unary potential corresponds to the initial saliency score, while the pairwise potential maximize label agreement between adjacent variables. It effectively increases the consistence of neighboring super-pixel regions. The pairwise energy function is formulated as

$$E(S) = \sum_{i \in \mathcal{V}} (S_i - \tilde{S}_i)^2 + \lambda \sum_{i \in \mathcal{V}, j \in \mathcal{N}_i} w_{ij} \cdot (S_i - S_j)^2. \tag{13}$$

The weight $w_{ij} = \exp(-\frac{d^2_{\text{color}}(c_i, c_j)}{2\sigma_w^2})$ measures the color similarity of the two adjacent variables, σ_w^2 controls the degree of similarity, and we empirically fix σ_w^2 to 0.1. The parameter λ in (13) is set to 20 to balance the initial saliency score and the value contributed from adjacent variables.

We modify the CRF framework in two aspects: (1) The CRF algorithm uses discrete random variable which may take a value from sets of labels (e.g., $\mathcal{L} = \{1, 0\}$), while we assign continuous value to each random variable, i.e., $S_i \in [0, 1]$. As such, the optimized continuous value represents more precise probability that a super-pixel belongs to the salient object. (2) In CRF framework, the MAP inference is done via graph cut based algorithm, while in our formulation, both terms in (13) are convex functions. It may be solved efficiently by employing analytic methods.

The solution to the pairwise energy function is

$$S^* = (I + 2\lambda(D - W))^{-1} \tilde{S}, \tag{14}$$

where I denotes an identity matrix, D is a diagonal matrix, where $d_{ii} = \sum_j w_{ij}$, W represents the weight matrix, $\tilde{S}$ is the vector of initial saliency values. Figure 5(c) presents the refined map. It indicates that we get more accurate saliency map with consistent foreground and background after the optimization procedure. In addition, the object boundary is well preserved.

5 Experimental evaluations

In this section, we evaluate the performance of the proposed method and compare it with 14 state-of-the-art salient object detection algorithms. These methods are chosen based on the following criteria:

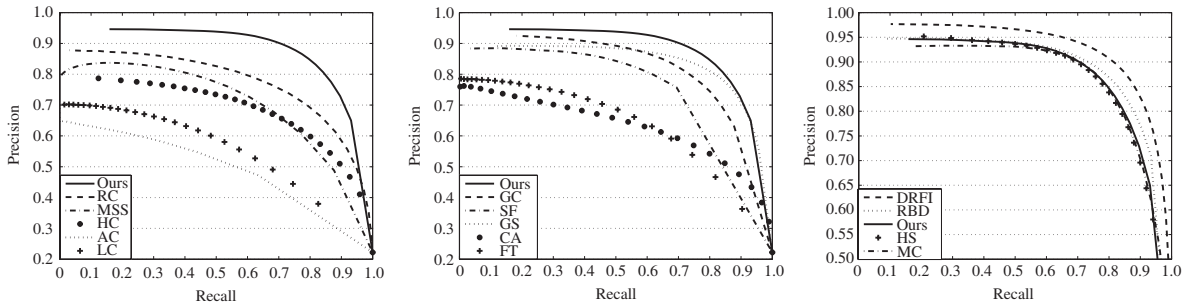


Figure 6 Precision recall curves comparison with fourteen state-of-the-art salient detection methods on dataset MSRA10K.

AC [17], CA [12], LC [37], FT [3] and MSS [38] are pixel contrast based methods, HC [4], RC [4], GC [32] and SF [18] adopt global regional contrast, MC [39] evaluates saliency via graph construction, GS [13] and RBD [20] use background prior, HS [14] detects saliency in a hierarchical manner, and DRFI [6] adopts a supervised learning method to classify salient and background regions. It is worth noting that DRFI and RBD are the top 2 approaches for salient object detection in the recent benchmark conducted by Ali Borji et al. [40].

5.1 Evaluation datasets

We have evaluated the results of the proposed method on three well known datasets: (1) MSRA10K[1)], (2) DUT-OMRON [34] and (3) SED2 [41]. They were also selected as standard benchmark datasets in [40]. Besides, all the three datasets have accurate pixel-wise binary masks.

MSRA10K: It is a challenging dataset containing 10000 images with complex scenarios. Note that it covers the whole 1000 images of the widely accepted ASD [3] dataset.

DUT-OMRON: It contains 5168 images selected from more than 140000 images. These images generally contain cluttered background and are annotated with both bounding boxes and pixel-wise binary masks.

SED2: It is a sub-dataset of SED [41] and has 100 images containing two salient objects. Objects of this dataset is annotated manually by three users. It is selected to evaluate the performance of the proposed method with multi-objects images.

5.2 Comparison with state-of-the-art methods

Three evaluation metrics are adopted to assess the performances of all the 15 salient object detection approaches.

Precision-recall (PR) curve: In the first evaluation, we compare different methods using precision-recall measure [40]. Precision is calculated as the percentage of pixels that are correctly assigned salient, while recall measures the percentage of detected salient pixels to ground truth masks. we first segment the normalized saliency maps using the fixed threshold that varies from [0:0.05:1] and then compare the obtained binary saliency maps to ground truth masks to obtain the precision-recall curve.

Figure 6 presents the fixed threshold precision-recall curves on dataset MSRA10K. It indicates that our result is comparable to RBD [20], MC [39] and HS [14]. Among them, RBD and MC is similar to our algorithm in graph construction, HS is similar to ours in that both methods utilize a multi-layer approach to analyze saliency cues. It is worth noting that DRFI, which model saliency using a supervised learning framework, outperforms the alternative data-driven methods in precision. However, this kind of methods needs exhaustive training on large-scale dataset, limiting its usability in many potential applications. Besides, our method is more accurate than all other approaches. Figures 7 and 8 present the results on DUT-OMRON and SED2 respectively. Both of them show similar results to MSRA10K and demonstrate the robustness of the proposed method over different kinds of datasets.

1) http://mmcheng.net/gsal/.

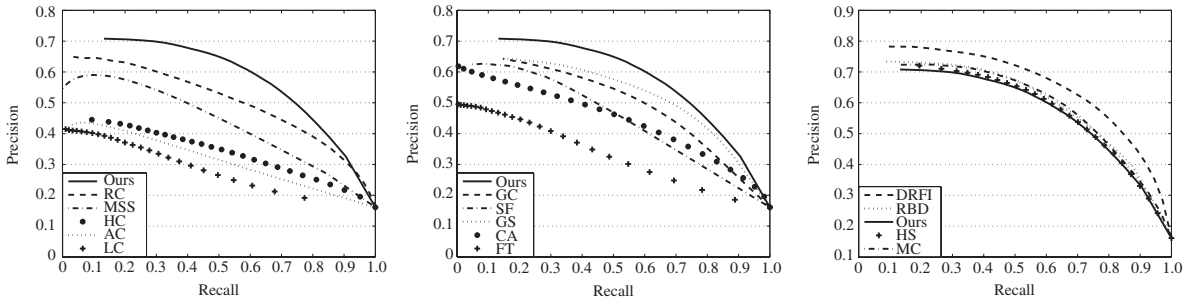


Figure 7 Precision recall curves comparison with fourteen state-of-the-art salient detection methods on dataset DUT-OMRON [34].

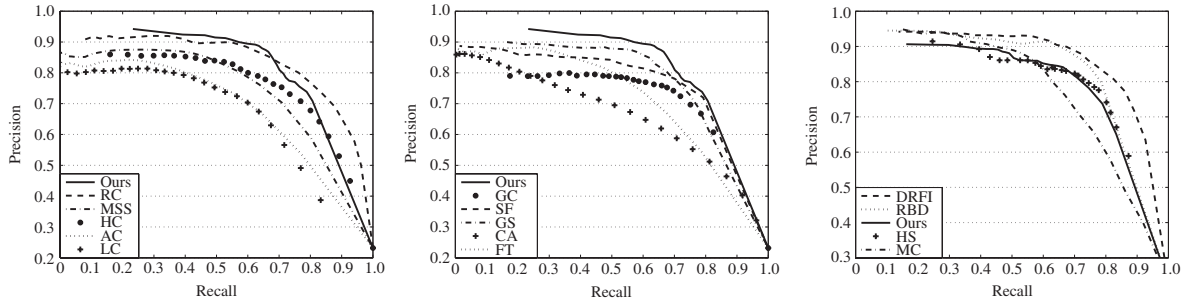


Figure 8 Precision recall curves comparison with fourteen state-of-the-art salient detection methods on dataset SED2 [41].

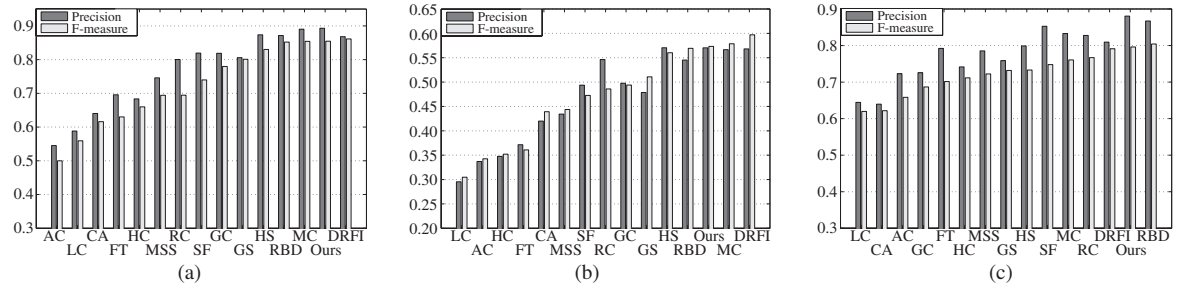


Figure 9 Adaptive threshold precision and F-measure comparison of saliency maps over three databases. (a) MSRA10K; (b) DUT-OMRON [34]; (c) SED2 [41].

F-Measure: In the second evaluation, we binarize the saliency map using an image-independent adaptive threshold as in [3,4,14]. We set the threshold to twice the mean saliency value in all experiments. Neither precision or recall can comprehensively evaluate the performance of saliency maps, thus we also calculate F-Measure. It is formulated as the weighted harmonic average of precision and recall:

$$F_\beta = \frac{(1+\beta^2)\cdot \text{Precision}\cdot \text{Recall}}{\beta^2\cdot \text{Precision} + \text{Recall}}. \tag{15}$$

We set $\beta^2 = 0.3$ as in [3, 4, 14], putting more weight to precision than recall. Figure 9 illustrates the precision and F-Measure results over three datasets. We can notice that the proposed method achieves comparable F-Measure performance to the top benchmark approaches (i.e., DRFI, RBD and MC) and consistently exceeds others. Besides, our approach obtains the best performance on the precision after we perform the threshold segmentation. This high precision means that the proposed approach could effectively depress the background regions when compared to alternate ones (see the visual comparison results in Figure 10).

Mean absolute error: Although being commonly used, precision-recall curve is limited in that neither the precision or recall measure considers the true negative counts, as pointed in [18, 32]. In the third evaluation, we calculate the mean absolute error (MAE). It is defined as the difference between the

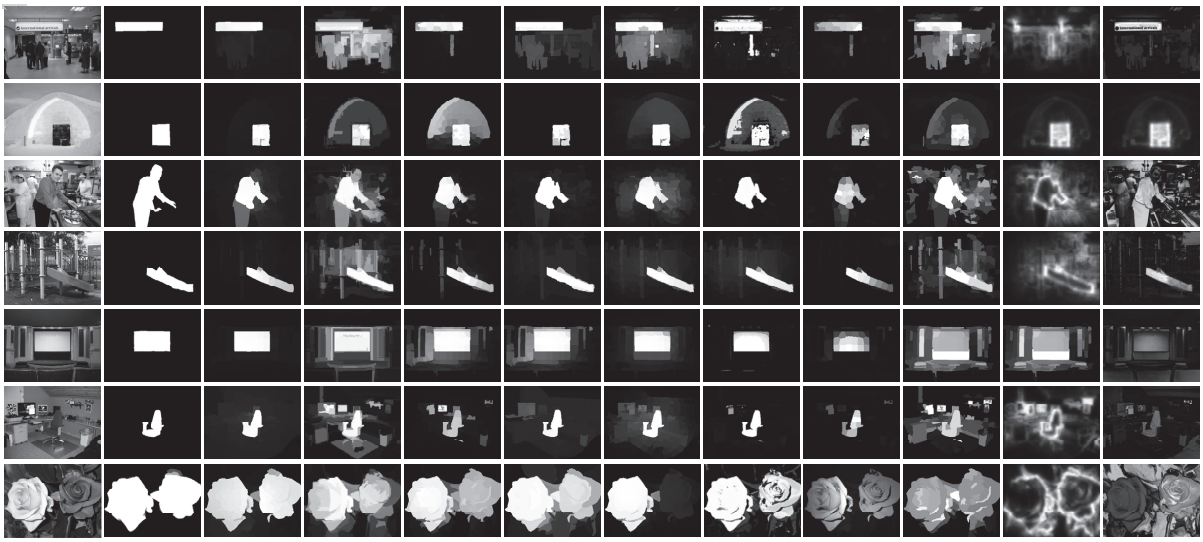


Figure 10 Qualitative comparison of saliency maps over complex scenes with cluttered background. We make comparison with discriminative region feature integration (DRFI), robust background detection (RBD), hierarchical saliency (HS), markov chain saliency (MC), global cues saliency (GC), filter based saliency (SF), geodesic saliency (GS), context-aware (CA), and frequency tuned saliency (FT). (a) Inputs; (b) GT; (c) ours; (d) DRFI; (e) RBD; (f) HS; (g) MC; (h) GC; (i) SF; (j) GS; (k) CA; (l) FT.

Table 1 MAE: Mean absolute error

Datasets	LC	AC	FT	CA	MSS	HC	RC	SF	HS	MC	GC	GS	DRFI	RBD	Ours
MSRA10K	0.233	0.227	0.235	0.237	0.203	0.215	0.252	0.175	0.149	0.145	0.139	0.147	0.118	0.108	0.126
DUT-OMRON	0.246	0.190	0.250	0.254	0.177	0.310	0.290	0.183	0.227	0.186	0.197	0.173	0.155	0.144	0.175
SED2	0.204	0.206	0.206	0.229	0.192	0.193	0.196	0.180	0.157	0.162	0.185	0.153	0.130	0.130	0.148

original saliency map S and its ground truth mask GT:

$$\mathrm{MAE} = \frac{1}{W \cdot H} \sum_{x=1}^{W} \sum_{y=1}^{H} |S(x,y) - \mathrm{GT}(x,y)|, \tag{16}$$

where W and H are the image width and image height respectively. We present the results in Table 1, which shows that the proposed method achieves top 4 performance over all the three datasets. It demonstrates that our approach could assign relatively precise value to the large-scale background region.

Qualitative comparisons: Figure 10 gives the visual comparison of the proposed method and the state-of-the-art algorithms. It indicates that our saliency maps accurately highlight the entire object and match the annotated ground truth masks well. When processing challenging images with complex background or multi-objects in the foreground, our proposed method can still produce high-quality saliency maps that consistently outperform state-of-the-art algorithms.

5.3 Validation of saliency measures

Our method uses various saliency cues and optimization measure. To analyze their respective significance in overall saliency estimation, we also evaluate each individual component on the ASD [3] dataset. The precision-recall results of different procedures in the proposed method are reported in Figure 11. We observe that region compactness cue and objectness-guided global distinctness cue have already achieved high-precision detection performance, and the two-layer saliency structure results achieve better detection accuracy than the individual component. It demonstrates that aggregating different saliency cues could contribute to the performance of saliency evaluation. Furthermore, the graph regularization measure further refines the results of the two-layer integration.

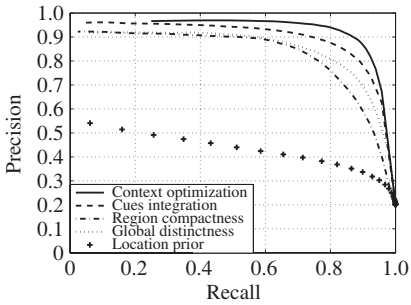


Figure 11 Validation of the proposed saliency measures. It indicates that individual compactness and distinctness measure already performs well, and the two-layer structure and graph regularization measure significantly improves detection accuracy.

Table 2 Comparison of the average runtime

Method	AC	FT	MSS	RC	SF	GS	GC	HS	MC	DRFI	RBD	Ours
Time (s)	0.15	0.09	0.088	0.15	0.16	0.35	0.06	0.6	0.53	1.17	0.39	0.19
Code	C++	C++	C++	C++	C++	Matlab	C++	EXE	Matlab	C++	Matlab	C++

Figure 12 Example images for which the proposed algorithm fail to produce good saliency maps. For each example, the left is the input image, the middle is saliency map and the right is the ground truth.

5.4 Computational efficiency

The computational efficiency comparison of the competitive algorithms is listed on Table 2. The time is evaluated as the average running time producing 10000 saliency maps on MSRA10K dataset on a laptop with 2.4 GHz Intel CPU and 4 GB RAM. The proposed method keeps low running time, while producing high-accuracy saliency maps. The most time-consuming part of our method is the image decomposition section (about 50%), which guarantees the evaluation efficiency in subsequent initial saliency value assigned part (about 40%) and saliency optimization part (about 10%).

5.5 Limitation

Saliency algorithms based on saliency cues which merely derives from color may not always generate accurate estimation. If there happens the salient regions share similar color with background (see Figure 12(a)), or the salient regions are cluttered (see Figure 12 (b) and (c)), the proposed method only highlights part of the salient regions, failing to assign consistent saliency value to the whole object. We would investigate alternate pattern cues (e.g., texture) and integrate them into our framework for future research. It is believed to dramatically benefit the saliency estimation, especially when detecting objects of irregular shape in natural scenes.

6 Conclusion

In this study, we propose a novel region-contrast based saliency estimation method. This method is realized by integrating three high-level saliency measures, then employing a pairwise energy minimization graphical model. It can generate saliency maps that uniformly highlight the salient regions and effectively suppress the background regions. We evaluate the proposed method extensively on three benchmark datasets and make comparison with 14 state-of-the-art algorithms. Experimental results varify the detec-

tion accuracy and efficiency of our method. Saliency map could be smoothed via weighted guided image filter in [42] with the input image as the guidance image. As such, the structure of the input image can be transferred to the saliency map better. This topic will be studied in our future research.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61573048, 51475017), Beijing Municipal Natural Science Foundation (Grant No. 4142033), and International Scientific and Technological Cooperation Projects of China (Grant No. 2015DFG12650).

Conflict of interest The authors declare that they have no conflict of interest.

References

1 Desimone R, Duncan J. Neural mechanisms of selective visual attention. Annu Rev Neurosci, 1995, 18: 193–222
2 Treisman A M, Gelade G. A feature-integration theory of attention. Cog Psychol, 1980, 12: 97–136
3 Achanta R, Hemami S, Estrada F, et al. Frequency-tuned salient region detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Miami, 2009. 1597–1604
4 Cheng M, Mitra N J, Huang X, et al. Global contrast based salient region detection. IEEE Trans Patt Anal Mach Intell, 2015, 37: 569–582
5 Yang C, Zhang L, Lu H, et al. Saliency detection via graph-based manifold ranking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Portland, 2013. 3166–3173
6 Jiang H, Wang J, Yuan Z, et al. Salient object detection: a discriminative regional feature integration approach. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Portland, 2013. 2083–2090
7 Liu Y, Li X Q, Wang L, et al. Interpolation-tuned salient region detection. Sci China Inf Sci, 2014, 57: 012104
8 Donoser M, Urschler M, Hirzer M, et al. Saliency driven total variation segmentation. In: Proceedings of IEEE International Conference on Computer Vision, Kyoto, 2009. 817–824
9 Hiremath P S, Pujari J. Content based image retrieval using color boosted salient points and shape features of an image. Int J Image Process, 2008, 2: 10–17
10 Feng J, Ma L, Bi F K, et al. A coarse-to-fine image registration method based on visual attention model. Sci China Inf Sci, 2014, 57: 122302
11 Marchesotti L, Cifarelli C, Csurka G. A framework for visual saliency detection with applications to image thumbnailing. In: Proceedings of IEEE International Conference on Computer Vision, Kyoto, 2009. 2232–2239
12 Goferman S, Zelnik-Manor L, Tal A. Context-aware saliency detection. IEEE Trans Patt Anal Mach Intell, 2012, 34: 1915–1926
13 Wei Y, Wen F, Zhu W, et al. Geodesic saliency using background priors. In: Proceedings of European Conference on Computer Vision, Florence, 2012. 29–42
14 Yan Q, Xu L, Shi J, et al. Hierarchical saliency detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Portland, 2013. 1155–1162
15 Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Patt Anal Mach Intell, 1998, 11: 1254–1259
16 Alexe B, Deselaers T, Ferrari V. What is an object? In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, 2010. 73–80
17 Achanta R, Estrada F, Wils P, et al. Salient region detection and segmentation. In: Proceedings of 6th International Conference on Computer Vision Systems, Santorini, 2008. 66–75
18 Perazzi F, Krahenbuhl P, Pritch Y, et al. Saliency filters: contrast based filtering for salient region detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Providence, 2012. 733–740
19 Han J, He S, Qian X, et al. An object-oriented visual saliency detection framework based on sparse coding representations. IEEE Trans Circ Syst Video Technol, 2013, 23: 2009–2021
20 Zhu W, Liang S, Wei Y, et al. Saliency optimization from robust background detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Columbus, 2014. 2814–2821
21 Han J, Zhang D, Hu X, et al. Background prior based salient object detection via deep reconstruction residual. IEEE Trans Circ Syst Video Technol, 2015, 25: 1309–1321
22 Ma L, Chen L, Zhang X J, et al. A waterborne salient ship detection method on SAR imagery. Sci China Inf Sci, 2015, 58: 089301
23 Shen X, Wu Y. A unified approach to salient object detection via low rank matrix recovery. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Providence, 2012. 853–860
24 Liu R, Cao J, Lin Z, et al. Adaptive partial differential equation learning for visual saliency detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Columbus, 2014. 3866–3873
25 Chang K Y, Liu T L, Chen H T, et al. Fusing generic objectness and visual saliency for salient object detection. In: Proceedings of IEEE International Conference on Computer Vision, Barcelona, 2011. 914–921
26 Li Y, Hou X, Koch C, et al. The secrets of salient object segmentation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Columbus, 2014. 280–287
27 Khuwuthyakorn P, Robles-Kelly A, Zhou J. Object of interest detection by saliency learning. In: Proceedings of

European Conference on Computer Vision, Heraklion, 2010. 636–649
28 Liu T, Yuan Z, Sun J, et al. Learning to detect a salient object. IEEE Trans Patt Anal Mach Intell, 2011, 33: 353–367
29 Kou F, Li Z, Wen C, et al. Perceptual based content adaptive L_0 smoothing. In: Proceedings of 14th Pacific-Rim Conference on Multimedia, Nanjing, 2013. 299–307
30 Achanta R, Shaji A, Smith K, et al. Slic superpixels. EPFL-REPORT-149300. 2010
31 Gopalakrishnan V, Hu Y, Rajan D. Salient region detection by modeling distributions of color and orientation. IEEE Trans Multimedia, 2009, 11: 892–905
32 Cheng M M, Warrell J, Lin W Y, et al. Efficient salient region detection with soft image abstraction. In: Proceedings of IEEE International Conference on Computer Vision, Sydney, 2013. 1529–1536
33 Margolin R, Tal A, Zelnik-Manor L. What makes a patch distinct? In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Portland, 2013. 1139–1146
34 Yang C, Zhang L, Lu H. Graph-regularized saliency detection with convex-hull-based center prior. IEEE Signal Process Lett, 2013, 20: 637–640
35 Xu L, Li H, Zeng L, et al. Saliency detection using joint spatial-color constraint and multi-scale segmentation. J Vis Commun Image Represent, 2013, 24: 465–476
36 Lafferty J, McCallum A, Pereira F C N. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proceedings of the 18th International Conference on Machine Learning. San Francisco: Morgan Kaufmann Publishers Inc., 2001. 282–289
37 Zhai Y, Shah M. Visual attention detection in video sequences using spatiotemporal cues. In: Proceedings of the 14th Annual ACM International Conference on Multimedia. New York: ACM, 2006. 815–824
38 Achanta R, Süsstrunk S. Saliency detection using maximum symmetric surround. In: Proceedings of IEEE International Conference on Image Processing, Hong Kong, 2010. 2653–2656
39 Jiang B, Zhang L, Lu H, et al. Saliency detection via absorbing markov chain. In: Proceedings of IEEE International Conference on Computer Vision, Sydney, 2013. 1665–1672
40 Borji A, Cheng M M, Jiang H, et al. Salient object detection: a benchmark. ArXiv e-prints, 2015
41 Alpert S, Galun M, Brandt A, et al. Image segmentation by probabilistic bottom-up aggregation and cue integration. IEEE Trans Patt Anal Mach Intell, 2012, 34: 315–327
42 Li Z, Zheng J, Zhu Z, et al. Weighted guided image filtering. IEEE Trans Image Process, 2015, 24: 120–129