Research Article

Salient Region Detection by Fusing Foreground and Background Cues Extracted from Single Image

Qiangqiang Zhou,¹ Weidong Zhao,¹ Lin Zhang,² and Zhicheng Wang¹

¹Department of Electronics & Information, Tongji University, Shanghai, China

²Department of Software Engineering, Tongji University, Shanghai, China

Correspondence should be addressed to Zhicheng Wang; zhichengwang@tongji.edu.cn

Received 17 May 2016; Revised 21 August 2016; Accepted 29 August 2016

Academic Editor: Erik Cuevas

Copyright © 2016 Qiangqiang Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Saliency detection is an important preprocessing step in many application fields such as computer vision, robotics, and graphics to reduce computational cost by focusing on significant positions and neglecting the nonsignificant in the scene. Different from most previous methods which mainly utilize the contrast of low-level features, various feature maps are fused in a simple linear weighting form. In this paper, we propose a novel salient object detection algorithm which takes both background and foreground cues into consideration and integrate a bottom-up coarse salient regions extraction and a top-down background measure via boundary labels propagation into a unified optimization framework to acquire a refined saliency detection result. Wherein the coarse saliency map is also fused by three components, the first is local contrast map which is in more accordance with the psychological law, the second is global frequency prior map, and the third is global color distribution map. During the formation of background map, first we construct an affinity matrix and select some nodes which lie on border as labels to represent the background and then carry out a propagation to generate the regional background map. The evaluation of the proposed model has been implemented on four datasets. As demonstrated in the experiments, our proposed method outperforms most existing saliency detection models with a robust performance.

1. Introduction

Many cognitive psychology researches have shown that given a visual scene, human attention is directed to particular parts by visual selective mechanism and these parts are called salient regions [1]. In computer vision, salient region detection simulates the functionality of selective attention and localizes and tags the attention-grabbing regions or pixels in a digital image. Borji et al. [2] provided a more precise definition; they think that a salient region detection model should first detect the salient attention-grabbing objects in a context and then segment the whole objects. Usually, a generated saliency map is the output of the model and the intensity of each pixel in map means its probability of belonging to salient regions. According to this definition, we can know that this issue is essentially a figure/ground segmentation problem, and the goal is to only segment the salient foreground object from the background. However, it is slightly different from

the traditional image segmentation problem that aims to partition an image into perceptually coherent regions.

Saliency detection for image is an important preprocessing step in many areas such as computer vision, graphics, and robotics to reduce computational cost by focusing on salient regions and neglecting the nonsalient. The value of salient region detection models lies on their wide applications, for instance, object detection and recognition [3], image and video compression [4], thumb-nailing [5], image quality assessment [6], image segmentation [7], content-based image retrieval [8], and so on.

Viewed from the information processing perspective, the existing models of visual saliency detection can be categorized as two main kinds: bottom-up and top-down. Bottom-up methods [9–11], also known as stimuli-driven or task-independent models, mainly detect saliency based on low-level feature attributes (color, orientation, motion, etc.) without any prior knowledge. Top-down approaches [12–15]

are often task-driven or scene-dependent models learning through training process, which requires some specific prior knowledge or high-level information.

Most of bottom-up saliency methods are based on the intuitive assumptions-appearance contrast between object and background. Depending on the range where the contrast is computed, the bottom-up method can be further divided into local methods [16–19] and global methods. Local methods compute various contrast measures in a local neighborhood of the pixel/patch, such as self-information [16], edge contrast [17], center-surround discriminative power [18], and center-surround differences [17, 19].

As a pioneer, Itti et al. [19] proposed employing center-surround differences across multi-scale image features to implement image saliency detection. Hereafter, many extended approaches based on this local contrast method have been invented, including graph based (GB) visual saliency model by Harel et al. [20] and a model (AC) presented by Achanta et al. [21] which determine salient regions by exploiting low-level information of color and luminance. Goferman et al. [22] propose a context-aware (CA) algorithm for realizing three principles involving local low-level clues, global considerations, and visual organization rules to highlight salient objects along with their contexts. However these local contrast methods still have several issues to be solved; typically, the boundaries of the salient object can be found well, but it cannot propagate the saliency to the object interior and highlight it uniformly, as shown in Figure 1(c). Moreover not all unique regions based on local contrast are salient; a small region with large local contrast may be viewed as meaningless distractor by the human.

Despite the fact that local contrast conforms to the psychobiology principle, global method should also be taken into account to capture the holistic rarity in image, when one region is similar to its neighborhoods while standing out in the whole scene.

Global methods generally exploit the whole image to compute the saliency of each pixel/region. Moreover, they assign comparable saliency values across similar regions. Specifically, they suppress the features that frequently occur, while maintaining those that deviate from the norm. Some methods assume globally less frequent features are more salient and use frequency analysis in the spectral domain [23, 24]. Other methods compare individual pixel/patch to the rest in image and use the averaged appearance dissimilarity as the saliency measure [25], and the averaging is usually weighted by spatial distances between pixels/patches to consider a fact that salient parts are usually grouped together to form a salient object.

However, global method has its inherent disadvantages (illustrated in Figure 1(d)). When a foreground region is globally compared with the remainder of the image, its contrast with the background is less prominent and the salient object is unlikely to be uniformly highlighted. And also most global methods rely mainly on color uniqueness in terms of global statistics, which can be insufficient to deal with complex variations common in natural images. Moreover, these methods ignore the spatial relations between different parts, which can be vital for reliable saliency detection.

Even when combining the local method with the global one, some critical problems remain to be addressed, such as objects containing salient small-scale patterns or backgrounds containing complex patterns, as shown in Figure 1(e). In the figure, the yellow flowers on the grass are likewise highlighted by the method mentioned in [17]; however, they are actually part of the background when viewing the picture as a whole.

Comparatively speaking, bottom-up saliency detection models based on local or global computing methods are usually more efficient for detecting fine details rather than global shape information. Top-down models, on the other hand, are typically used to detect objects of specific sizes and categories based on more representative features from training samples. In addition, these models can leverage high-level prior knowledge. For example, most background regions can be easily connected to the image boundaries mentioned in [26]. Nevertheless, the result tends to be coarse with fewer details. In terms of computational complexity, bottom-up approaches are often more efficient than top-down ones.

Because both bottom-up and top-down models have drawbacks, it is natural to integrate them to more effectively and efficiently detect salient objects. Reference [27] proposed a new computational visual attention model by combining bottom-up and top-down mechanisms for man-made object detection in scenes. This model shows that the statistical characteristics of orientation features can be used as top-down clues to help determining the location for salient objects in natural scenes. Reference [28] invented a computational model that performs such a combination in a principled way. The system learns an optimal representation of the influences of task and context and thereby constructs a biased saliency map representing the top-down information. This map is combined with bottom-up saliency maps in a process progressing over time as a function over the input. But this method is not universal and it is only applied to the study of visual fixation shift. Reference [29] proposed a novel salient region model based on the bottom-up and top-down mechanisms: the color contrast and orientation contrast are adopted to calculate the bottom-up feature maps, while the top-down cue of depth-from-focus from the same single image is used to guide the generation of final salient regions, since depth-from-focus reflects the photographer's preference and knowledge of the task. Liu et al. [17] formulate salient object detection as a binary labeling problem and developed a conditional random field method to effectively combine bottom-up and top-down approaches. At the same time, a set of novel features including multi-scale contrast, center-surround histogram, and color spatial distribution are put forward to describe a salient object locally, regionally, and globally.

However, the most existing integrated methods including those mentioned above adopt a single similar linear weight to combine the feature maps computed by bottom-up or top-down approaches. Thus, the single similar linear weight loses universality and the performance is significantly decreased when it is applied to various types of images. In addition, existing integration models lack effectiveness and sufficient

FIGURE 1: Comparing different approaches to saliency.

flexibility because several parameters must be empirically specified.

In this paper, a method is proposed to address the above problems. In Figure 2, the proposed saliency detection algorithm is outlined. The contributions of our work include the following.

- (1) A new local contrast map measurement method provides greater alignment with physiological law.
- (2) A bottom-up coarse salient region extraction method combines the local method with the global method. The global method is derived based on two kinds of prior information.
- (3) A novel top-down background measure is via boundary label propagation. We first construct an affinity matrix. Then, we select and label nodes existing on the border to represent the background. We then carry out a propagation to generate the regional background map.
- (4) Unlike the conventional, commonly used linear weighted combination, we present a unified approach to integrating bottom-up, low-level features and top-down, high-level priors for saliency optimization and refinement.

Figure 3 depicts examples of the proposed method. To verify the effectiveness of our proposed approach, we performed extensive comparisons on a benchmark dataset. Performance was evaluated in terms of precision-recall, F-measure, and mean absolute error (MAE). We compared our method with ten state-of-the-art techniques. Both quantitative and qualitative experimental results demonstrate that our proposed method achieves stable and excellent performance against other methods.

2. Coarse Salient Regions Extraction

In this section, we formulate the computations relating to the three components used in the proposed coarse salient region extraction method: the local contrast map, focus prior map, and global color prior map. After obtaining these three maps, we then combine them to enable coarse salient region extraction. However, it should be noted that, before

computing the local contrast map, we must first employ the structure extraction algorithm [30] for preprocessing.

2.1. Local Contrast Map Measure

2.1.1. Structure Extraction. Some physiology experiments [31] have demonstrated that the human visual system is more sensitive to patterns in middle frequencies than those in high frequencies. Figure 4, respectively, depicts input images and texture-suppressed images. At first glance, viewers are attracted by the flowers in the images, while ignoring the minor variations of leaves surrounding them. However, the effects of such minor variations within textures may accumulate and cause some difficulties in the detection of salient objects. Therefore, we adopt the structure extraction algorithm proposed in [30] to suppress these small variances in textures. This algorithm smoothens the local gradients in textures, preserves the global structures of the objects, and diminishes insignificant details, thereby producing more accurate boundary information. The objective function is expressed as

$$\arg \min_S \sum_p (S_p - I_p)^2 + \lambda \cdot \left(\frac{D_x(p)}{L_x(p) + \varepsilon} + \frac{D_y(p)}{L_y(p) + \varepsilon} \right), \quad (1)$$

where I is a given image, which could be the luminance (or log luminance) channel, and p indexes pixels. S is the extracted structure image. The data term $(S_p - I_p)^2$ serves to make the extracted structures similar to those in given image. The second term is a regularizer named as *relative total variation* which is effective at emphasizing the main structures. Different values of the parameter λ in (1) output images with various smoothness level, and ε is a small positive number to avoid division by zero. The larger λ is, the smoother the image is. $D_x(p)$ and $D_y(p)$ are the sum of the absolute spatial difference in the x and y directions weighted by a Gaussian function within a window for pixel p . $L_x(p)$ and $L_y(p)$ are the modulo of the directional spatial difference sum in the x and y directions weighted by a Gaussian function within a window for pixel p .

FIGURE 2: The framework of our proposed saliency detection algorithm.

FIGURE 3: Examples of the proposed methods.

FIGURE 4: Input images and their texture-suppressed images.

After this structure extraction, the texture-suppressed image is then segmented into N superpixels by the simple linear iterative clustering (SLIC) algorithm [32] for postprocessing.

2.1.2. Weber Local Descriptor. Inspired by Weber's law, the Weber local descriptor (WLD) [33] is based on a physiological law [34]. Weber's law implies that the change of a stimulus (such as lighting) that is minimally noticeable is a constant ratio of the original stimulus. If the change is smaller than this constant ratio of the original stimulus, the human visual system does not recognize it as a valid signal; rather, it is recognized as background noise. Accordingly, WLD extracts features from the image by simulating the way a viewer perceives the surroundings. It consists of two components—differential excitation and orientation. The differential excitation of a given pixel is computed based on the ratio between two items. One is the relative intensity difference of a current pixel against its neighbors (e.g., a 3×3 region); the other is the intensity of the current pixel. After obtaining the differential excitation, we can attempt to derive the local salient patterns from the input image.

Based on the implication of Weber's law, we must compute the two constituent parts of the WLD feature for each pixel in the input image: differential excitation and gradient orientation. In addition, four filtering structures f_{00} , f_{01} and f_{10} , f_{11} are additionally required in the calculation process as follows (see Figure 5).

Suppose the differences between the center point p_c and its neighbors p_i ($i = 0, 1, \dots, n-1$) are the output of filter f_{00} , which is calculated as

$$d_f^{00} = \sum_{i=0}^{n-1} (\Delta p_i) = \sum_{i=0}^{n-1} (p_i - p_c). \quad (2)$$

Then the differential excitation $\xi(p_c)$ on pixel p_c is computed as

$$\xi(p_c) = \arctan \left[\frac{d_f^{00}}{d_f^{01}} \right] = \arctan \left[\sum_{i=0}^{n-1} \left(\frac{p_i - p_c}{p_c} \right) \right]. \quad (3)$$

Here, the value $\xi(p_c)$ may take a minus value when the intensities of the neighbors are smaller than that of the current pixel. That is, positive $\xi(p_c)$ means that the surroundings are lighter than the current pixel, while negative $\xi(p_c)$ simulates the darker surroundings.

Similarly, we compute the orientation component of WLD as the gradient orientation described in [33], which is defined as

$$\varphi(p_c) = \arctan \left(\frac{p_f^{11}}{p_f^{10}} \right), \quad (4)$$

where p_f^{10} and p_f^{11} are the outputs of the filters f_{10} and f_{11} .

As described in [33], WLD provides several advantages. For one, it can detect edges that elegantly conform to the subjective criterion because it relies on the perceived luminance difference. Furthermore, WLD performs a powerful feature extraction method for textures and is robust to noise effects in a given image. Most importantly, WLD is based on physiological law. It extracts features from an image by simulating a human's perception of his/her surroundings.

2.1.3. Local Contrast Map. In this section, we define a function to measure the local contrast map based on the following three properties:

- (1) The higher the degree to which a superpixel contrasts with its neighbors, the greater its saliency value.
- (2) A superpixel has a much higher likelihood of being salient if it is closer to the image center.
- (3) A region is more likely to pertain to the background if it has a large number of pixels on the image boundary. Accordingly, we calculate the integrity of a region with consideration of the amount of pixels located on the boundary.

First, segment a given image into K superpixels. For a superpixel s_i and its corresponding neighborhood $\{n_j\}$, $i = 1, \dots, K_i$, where K_i represents the number of its neighbors. Then we can define the saliency computation of s_i as

$$S(s_i) = \left(\sum_{j=1}^{K_i} \omega_{ij} f(d(s_i, n_j)) \right) \times t(x, y) \times h(u), \quad (5)$$

where $d(s_i, n_j) = \max(\text{col Dist}(s_i, n_j), \text{wld Dist}(s_i, n_j))$ denotes the feature distance between superpixel s_i and n_j , where $\text{col Dist}(s_i, n_j)$ and $\text{wld Dist}(s_i, n_j)$, respectively, represent the χ^2 distance of LAB color histograms and Weber local descriptor (WLD) histograms [33]. In addition, $d(s_i, n_j)$ is the histogram Euclidean distance, and variable ω_{ij} is a ratio between neighbor region n_j and the total area of its neighborhood $\{n_j\}$.

Function $f(\phi)$ guarantees that the output will be positive. Thus, to highlight the salient regions, we use

$$f(\phi) = -\log(1 - \phi). \quad (6)$$

+1	+1	+1	0	0	0		-1				
+1	-8	+1	0	+1	0				+1		-1
+1	+1	+1	0	0	0		+1				
f_{00}			f_{01}			f_{10}			f_{11}		

FIGURE 5: Four filtering structures of WLD.

In (5), x and y denote the average position of pixels in the superpixel and they are normalized to $[0, 1]$. Moreover, $t(x, y)$ is a normalized distance between the superpixel center s_i and the image center (x_0, y_0) , which can be expressed by

$$t(x, y) = \exp\left(-\frac{(x - x_0)^2}{(2\sigma_x^2)} - \frac{(y - y_0)^2}{(2\sigma_y^2)}\right), \quad (7)$$

where σ_x is set as $1/3$ of the image width and σ_y is set as $1/3$ of the image height. Therefore, the superpixel that is closer to the center of the image will attain a greater weight.

We define the integrity $h(u)$ in (5) as

$$h(u) = \begin{cases} \exp\left(-\frac{u}{\lambda \times E}\right), & \frac{u}{E} \leq \eta, \\ 0, & \text{otherwise,} \end{cases} \quad (8)$$

where u represents the number of pixels on the image boundary contained by superpixel s_i . In addition, E denotes the total number of pixels on the image boundary, λ is an adjustment for tuning the strength of its impact, and η is a threshold ($\lambda = 0.05$, $\eta = 0.07$). A superpixel with a larger u means it is unlikely to be an integral object. If $u = 0$, it indicates that a region is not on the boundaries, and $h(u) = 1$. Otherwise, $h(u)$ is a value within $[0, 1]$.

2.2. Frequency Prior. The arguments proposed by Achanta et al. [23] indicate that the mechanism of the human visual system detects the salient objects in a scene which can be well modeled by integrating band-pass filtering responses from the CIE $L^*a^*b^*$ color channels. By means of the filter of DoG (*Difference of Gaussian*), Achanta et al. achieve the desired results of band-pass filter. Inspired by their works, in this paper, we also apply the band-pass filter to salient object detection. However, in view of the following considerations we choose the log-Gabor filter instead of DoG.

Firstly, Field [35] suggests that natural images are better coded by filters that have Gaussian transfer functions when viewed on the scale of logarithmic frequency. On the linear frequency scale the log-Gabor function $g(x)$ ($x = (x, y) \in \mathbb{R}^2$) has a transfer function whose form can be expressed as

$$G(f) = \exp\left(\frac{-(\log(f/f_0))^2}{2(\log(\delta/f_0))^2}\right), \quad (9)$$

where f_0 is the filter's center frequency and δ/f_0 controls the filter's bandwidth. By definition, log-Gabor functions always

have no DC component and can construct a log-Gabor filter with an arbitrarily bandwidth. Secondly, the transfer function of the log-Gabor function has an extended tail at high frequency end which makes it more capable to encode the images more efficiently than other ordinary Gabor functions. $g(x)$ cannot be analytically expressed due to the singularity in the log function at the origin. Instead, $g(x)$ can only be approximately obtained by performing a numerical inverse Fourier transform to $G(f)$. An illustration of a 2D log-Gabor filter in frequency domain ($f_0 = 1/6$, $\delta/f_0 = 0.3$) is shown in Figure 6.

Suppose an image $\{I(x) \mid x \in \Omega\}$, where Ω defines a spatial domain in RGB color space and $I(x)$ is a triple vector containing three intensities values (R, G, B) at position x . Similar to [23], we can obtain its frequency prior map $S_a(x)$ which is modeled by band-pass filtering as follows: at first, the image $I(x)$ needs to be converted to an opponent color space such as CIE $L^*a^*b^*$ in which the three resulting channels are represented by $I_L(x)$, $I_a(x)$, and $I_b(x)$. Then we can defined $S_a(x)$ as

$$S_a(x) = \left((I_L * g)^2 + (I_a * g)^2 + (I_b * g)^2\right)^{1/2}(x), \quad (10)$$

where $*$ denotes a convolution operation. For more details refer to ours [36]. An example is shown in Figure 7.

2.3. Color Spatial Distribution. From Figure 8 we can observe that if one type of color is distributed in an image extensively, the more possible this color forms the background, and the less possible a salient object contains it. In other words, a specific color with a smaller spatial variance is more likely to be part of the salient object. And the goal of using color spatial distribution is to take the global color information in image into consideration.

One simple approach to characterize the spatial variance of the color is based on the Gaussian Mixture Model (GMM), which is one of the most statistically ordinary algorithms for clustering [17].

Generally, each pixel in the GMM can be denoted as a probability assigned to a color component.

$$p(c \mid I_x) = \frac{\omega_c N(I_x \mid \mu_c, \Sigma_c)}{\sum_c \omega_c N(I_x \mid \mu_c, \Sigma_c)}, \quad (11)$$

where ω_c , μ_c , and Σ_c are the weight, mean color, and covariance, respectively, and $N(I_x \mid \mu_c, \Sigma_c)$ is the multivariate normal distribution of the c th component.

FIGURE 6: A 2D log-Gabor filter in frequency domain.

FIGURE 7: An example about frequency prior map.

For each associated color component c , its corresponding horizontal variance $V_h(c)$ of the spatial position can be computed by

$$V_h(c) = \frac{1}{|X|_c} \sum_x p(c | I_x) \cdot |x_h - M_h(c)|^2, \quad (12)$$

where x_h is x -coordinate of the pixel x , $|X|_c$ is a normalizing factor and $|X|_c = \sum_x p(c | I_x)$, and $M_h(c)$ is the mean of the Gaussian component:

$$M_h(c) = \frac{1}{|X|_c} \sum_x p(c | I_x) \cdot x_h. \quad (13)$$

Similarly we can also define the vertical variance $V_v(c)$ and thus derived the spatial variance of a component c by combining the horizontal and vertical variances:

$$V(c) = V_h(c) + V_v(c). \quad (14)$$

Then employ a min-max approach to normalize this composite covariance to the range $[0, 1]$:

$$V(c) \leftarrow \left(V(c) - \frac{\min(V(c))}{\max(V(c))} - \min(V(c)) \right). \quad (15)$$

Finally, the color spatial distribution feature map $f_s(x, I)$ is therefore defined as a weighted sum of its color distribution

of each pixel, due to the assumption that a salient object tends to have a less widely distributed color:

$$f_s(I_x, I) \propto \sum_c p(c | I_x) \cdot (1 - V(c)). \quad (16)$$

The feature map $f_s(x, I)$ is also normalized to fall within $[0, 1]$. Figure 8 exemplified several images to demonstrate the color spatial distribution feature maps, from which we can observe that these salient objects are highlighted effectively by this global feature.

It is obvious that exploiting this global feature can detect salient objects with a high accuracy when the background in image is monotonous. However, in the case of complex and colorful background, the color spatial distribution map fails to distinguish the salient object in the image. Figure 9 demonstrates to us this undesired property.

2.4. Map Integrating. After generating these three maps—local contrast map (LCM), frequency prior (FP), and color spatial distribution (CSD)—we can then derive the coarse salient map, SM_c , through the following combination:

$$SM_c = \frac{(LCM + FP + CSD)}{3}, \quad (17)$$

where LCM, CPM, and FPM are all normalized to $[0, 1]$ and LCM is computed in a local region perspective. Meanwhile, FP and CSD are computed in a global perspective. They are complementary to each other.

FIGURE 8: Color spatial distribution map.

FIGURE 9: Three undesired examples.

3. Background Extraction Based on Boundary Label Propagation

3.1. Affinity Matrix Construction. In this section, we construct an affinity matrix based on boundary label propagation for the subsequent computation of background extraction. As discussed in Section 2.1.1, the input image must implement texture suppression. Then, it is segmented into N superpixels by the SLIC algorithm. We regard these generated superpixels as nodes or regions in this affinity matrix construction procedure. In addition, those regions on the image border form a boundary set denoted as $\mathbf{B}$.

Thereafter, we can define a superpixel-based graph $G = (V; E)$, where V is a set of nodes and E is a set of undirected edges weighted by an affinity matrix $\mathbf{W} = [w_{ij}]_{N \times N}$ to indicate the similarity between any pair of nodes. The degree matrix of

a given graph G is defined as $\mathbf{D} = \text{diag}\{d_{11}, \dots, d_{NN}\}$, where $d_{ii} = \sum_j w_{ij}$.

In this work, there are two adaptations to construct the affinity matrix entry by means of adding two constraints.

First, we connect each node to its 2-ring neighbors to fully employ the adjacency information of nearby nodes. This is because neighboring nodes are more likely to have a similar appearance and saliency (see Figure 10). That is, each node is not only connected to those neighboring it, but it is also connected to the nodes sharing common boundary with its neighboring node. By extending the scope of node connection, we effectively exploit the local smoothness cues.

Second, based on a background prior revealed in [26], there is a cue that the image boundary of most background and salient objects rarely touches the image boundary. This cue is more common than the prior of center bias that is

Input: $\mathbf{A}_{N \times N}$: Row-normalized affinity matrix L_0 : Set of given boundary nodes' labels $U = \{S \setminus L_0\}$: Set of unlabelled nodes $t = 0$ Initialize, set $f_t(s_i) = 1$ for $s_i \in L_0$ and $f_t(s_i) = 0$ for $s_i \in U$ **while** $VAR > threshold$ **do** **for** $s_i \in U$ **do**

$$f_{t+1}(s_i) = \sum_{j=1}^N a_{ij} f_t(s_j)$$

end for $t = t + 1$ $VAR = \text{var}(f_t, f_{t-1}, \dots, f_{t-\text{constant}})$ **end while** $\mathbf{BG} = \text{normalize}(f_t)$ $\mathbf{BG}(s_i) = \text{map}(\mathbf{BG})$ **Output:**Regional background probability weight map $\mathbf{BG}(s_i)$ generate by background labels

ALGORITHM 1: Label propagation via boundary nodes.

FIGURE 10: Graph model with superpixels as nodes (the purple and green lines represent connections to first- and second-order neighbors, respectively. The red line means that the all boundary nodes are connected among themselves).

often used. Thus, we enforce all the superpixels which lie on the four sides of image are connected; hence, any boundary node pair is handled as being adjacent (see Figure 10). This approach can reduce the geodesic distance of similar superpixels and further enhance the relationship among boundary nodes. Because we handle all boundary nodes as propagation labels in Algorithm 1, a tight connection among them can more effectively distinguish the background from the foreground.

Using the definitions on the graph and the constraints on the edges, we define the affinity entry w_{ij} of node i to a certain node j as

$$w_{ij} = \begin{cases} \exp\left(-\frac{\|c_i - c_j\|}{\sigma^2}\right), & j \in N(i) \text{ or } i, j \in B, \\ 0, & i = j \text{ or otherwise,} \end{cases} \quad (18)$$

where c_i, c_j denote the mean feature vectors of pixels included in the superpixels corresponding to the two nodes in the

CIELAB color space and σ is a constant for controlling the strength of the weight. Furthermore, $N(i)$ indicates the set of neighboring nodes existing in the 2-ring range of node i . According to (18), we can additionally obtain a $N \times N$ row-wise normalized color affinity matrix:

$$\mathbf{A} = \mathbf{D}^{-1} \cdot \mathbf{W}. \quad (19)$$

3.2. Propagation via Boundary Labels. In this section, we introduce a label propagation algorithm to compute the background probability weight of other unlabeled superpixels by setting the boundary nodes to serve as initial propagation labels.

Given a dataset $S = \{s_1, \dots, s_l, s_{l+1}, \dots, s_n\} \in \mathbf{R}^{m \times n}$, where m denotes the data dimension and the former l data points are labeled queries; the remaining data points can be calculated by propagating the information of similarity correlation with the queries. Let $\mathbf{f} : S \rightarrow [0, 1] \in \mathbf{R}^n$ denote a mapping function that assigns a propagating value f_i to each point s_i , to indicate the degree to which each data point is similar to the labels, and $\mathbf{f}$ can be regarded as a vector $\mathbf{f} = [f_1, \dots, f_n]^T$. The similarity measure $f(s_i)$ satisfies

$$f_{t+1}(s_i) = \sum_{j=1}^N a_{ij} f_t(s_j), \quad (20)$$

where t denotes the recursive times and a_{ij} is the affinity value in (19).

Suppose that the nodes in our build graph constitute the given dataset, S . During the process of recursive calculation, the similarity indications of query labels are set to 1, while the initial indication of unlabeled nodes is set to 0. For a given node, the similarity value $f(s_i)$ is learned through iteratively propagating the similarity measure of its neighbors, $f(s_j)$. Hence, its ultimate value of similarity to the labels is influenced by its surroundings.

FIGURE 11: Quantitative evaluations by (a) PR curves, (b) adaptive threshold, and (c) MAE on dataset MSRA10K.

Algorithm 1 outlines the entire procedure of label propagation via boundary nodes. Its output regional background probability weight map forms the extracted background. Its convergence condition constraint is found by determining whether the average variance of similarity measure $\mathbf{f}$ in the last 50 iterations (e.g., $constant = 49$) is lower than a given threshold. The saliency measures of N regions are converted into a pixel-wise map by $\text{map}(\cdot)$. This type of propagation process is similar to the proposal in [37]. This label propagation via boundary nodes works well in most cases, as shown in Figure 3(c).

4. Saliency Refinement

The extracted background map, BG, in Algorithm 1 and the coarse saliency map, SM_c , obtained in Section 2, are complementary and mutually reinforcing because they characterize the background and foreground, respectively. Nevertheless, they are still coarse and noisy. Thus, we propose a principled framework to integrate these two methods and generate the final refined saliency map, as described in Figure 2.

Commonly speaking, after extraction of multiple saliency cues according to different feature measures, a critical step is their combination strategy for integrating these feature maps. By far, there have been many kinds of framework for such computation, but most of them are based on the hypothesis that similar features compete for saliency, while different modalities contribute independently to the saliency detection [38]. In fact, there is a lot of psychological evidence to show that different features types have different contribution to the saliency, and not just a simple combined heuristically using weighted summation or multiplication. Thus many methods have been put forward to integrate different kinds of feature map for visual saliency recognition. As mentioned in [39], four combination methods are proposed—(1) simple normalized summation, (2) linear combination with learned weights, (3) global nonlinear normalization followed by summation, and (4) local nonlinear competition between salient locations followed by summation. Also [38], inspired by the role of sparse in human perception and the sparse property of human visual neurons, defined a feature sparse indicator that measures feature's contribution

FIGURE 12: Quantitative evaluations by (a) PR curves, (b) adaptive threshold, and (c) MAE on dataset ECSSD.

to saliency map. In addition, [40] proposed AdaBoost as the central computational module that takes into account feature selection, weight assignment, and integration in a principled and nonlinear learning framework. One of approaches most close to our idea was presented in [17]; a conditional random field is learned to effectively combine multiple visual cues for salient object detection through defining an energy function.

In this paper, the proposed framework combines the saliency cues derived from the foreground and background, respectively, and formulates the saliency detection as an optimization problem of the saliency values of all superpixels in the given image. It then achieves an optimal saliency map by minimizing our designed objective cost function, such that it, respectively, assigns value 1 to the object region and value 0 to the background region.

Our cost function consists of three different constraint terms and is defined as shown in the following formula:

$$\sum_{i=1}^N w_i^{\text{BG}} s_i^2 + \sum_{i=1}^N w_i^{\text{SM}_c} (s_i - 1)^2 + \sum_{i,j} w_{ij} (s_i - s_j)^2, \quad (21)$$

where s_i ($i = 1, \dots, N$) represents the final saliency value assigned to superpixel p_i after minimizing the cost.

The first term prompts superpixel p_i with a high background probability w_i^{BG} to obtain a small value approximated to 0. As mentioned in Section 3, w_i^{BG} has a high accuracy derived from our effective background detection. Accordingly, in the second term, $w_i^{\text{SM}_c}$ denotes the foreground weights associated with superpixel p_i , while large $w_i^{\text{SM}_c}$ boosts p_i to take values close to 1. After optimization, the foreground weight map is significantly improved on account of our proposed optimization framework.

The last term is a *smoothness* term that encourages continuous saliency values. For any pair of adjacent superpixels, such as p_i and p_j , the weight w_{ij} can be defined as

$$w_{ij} = \exp\left(-\frac{d_{\text{app}}^2(p_i, p_j)}{2\sigma^2}\right) + \tau. \quad (22)$$

It has a large value in the flat region; however, it has a small value on the region boundary. To minimize the influence

FIGURE 13: Quantitative evaluations by (a) PR curves, (b) adaptive threshold, and (c) MAE on dataset SED2.

of minor noise in both background and foreground terms and to regularize the optimization in cluttered regions, we empirically introduce a parameter τ and set it as 0.1, where the parameter $\sigma = 10$, and $d_{\text{app}}(p, q)$ denotes the Euclidean distance between their average colors in CIE $L * a * b$ color system.

Equation (21) consists of three squared error terms. The optimization problem of the saliency map can be solved by using the least squares method.

5. Experiments

In this section, we will compare our proposed method with 8 state-of-the-art salient object detection schemes on four typical available datasets.

5.1. Datasets and Evaluation Metrics

Datasets. Four benchmark datasets for evaluation include commonly used MSRA10K [17] (10000 images), Judd [41] (900 images), SED2 [42] (two-object set) containing 100 images, and ECSSD [43] (1000 images with texture background). The MSRA10K comprises per-pixel ground truth

annotation for 10000 MSRA images, each of which has an unambiguous salient object and the object region is accurately annotated with pixel-wise ground truth labeling. ECSSD is a dataset which includes 1000 semantically meaningful but structurally complex images for evaluation. Judd contains 900 images with multiple objects with high background clutter. SED2 is one of the two subsets of SED and has 100 images containing two salient objects. The reason to employ this dataset is to evaluate the accuracy of models on SED2 to verify whether saliency detection algorithms can work well on images containing two salient objects.

We compare our proposed method with 8 state-of-the-art salient region detection methods: GMR [44], LMLC [45], SVO [46], HS [43], CA [22], PCA [47], CB [48], and GR [49].

Evaluation Metrics. We use the precision, recall, F-measure (F_β), and mean absolute error (MAE) for our algorithm evaluation and comparison. These metrics are defined as follows:

$$\begin{aligned} \text{Precision} &= \frac{|B \cap \text{GT}|}{|B|}, \\ \text{Recall} &= \frac{|B \cap \text{GT}|}{|\text{GT}|}, \end{aligned} \quad (23)$$

FIGURE 14: Quantitative evaluations by (a) PR curves, (b) adaptive threshold, and (c) MAE on dataset Judd.

$$F_{\beta} = \frac{(1 + \beta^2) \text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}}, \quad (24)$$

$$\text{MAE} = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H |\text{Sal}(i, j) - \text{GT}(i, j)|. \quad (25)$$

In the above formulas, GT is the ground truth map, B denotes a binary mask generated by thresholding the obtained saliency map Sal via setting a threshold value, and $|\cdot|$ in (23) denotes the area (i.e., pixel number) of corresponding image region. Moreover, to acquire both high precision and high recall, we evaluate F-measure as (24), and the parameter β^2 is usually set as 0.3 to raise more importance for the precision value. Sal(i, j) in (25) is a saliency value in a pixel location (i, j), and W and H are the width and height of the saliency map and ground truth map, respectively.

In theory, a saliency map should agree with the ground truth map. Thus, the more similar the saliency map is to the ground truth, the more effective the generation of the saliency map algorithm is. Nevertheless, the precision-recall curves have limitations in this performance measure. We therefore additionally introduce MAE as a complementary metric. It

is a normalized value denoting the average per-pixel errors between the continuous saliency map, Sal, and the binary ground truth map, GT. MAE measures how consistent a saliency map is with the ground truth map. A good saliency detection method should achieve high precision-recall curves while maintaining low MAE.

Parameters Setup. In the experiment, all parameters are determined as follows: the number of segment superpixels is set as 600; the value of filter's center frequency $f_0 = 0.002$, and the parameter which controls the filter's bandwidth $\delta/f_0 = 6.2$ appears in Section 2.2; we set the control of color distance $\sigma_c^2 = 0.1$ in (18). The above parameters configuration achieves desired performance in our experiments.

5.2. Quantitative Comparisons. Here we use three standard measures—precision-recall (P-R) curve, F-measure, and MAE for evaluating saliency maps produced by different models on four datasets and making comparisons.

Fixed Threshold. For each method, we first bipartite the generating saliency map using a fixed threshold that is varied from 0 to 255. On each threshold, the pair of precision/recall scores

FIGURE 15: Qualitative comparisons of our saliency maps and recently developed approaches on MSRA10K dataset.

is computed by comparing the binarized map with the ground truth. These are finally combined to obtain the PR curves to evaluate the model performance in different states, as Figures 11(a), 12(a), 13(a), and 14(a) show. As we can see, compared with the all other approaches, our method shows the most consistent performance and ranks in top three best over 4 datasets. The performance of our technology is comparable to the outstanding mainstream methods including GMR, HS. The performances of LMLC and GR behave in an unstable manner on these four datasets and they perform well on the MSRA10K dataset, however with unsatisfied results on the other three datasets due to their weak background suppression ability. Our method significantly outperforms CB, CA, and LMLC on the datasets—ECSSD, Judd, and SED2. Besides, observing PR curves, our method achieves close performance to GMR and HS on MSRA10K and ECSSD datasets but outperforms GMR on SED2 dataset.

Adaptive threshold: in addition, adaptive threshold experiments were conducted to compute F-measure, where the binarizing threshold is defined as twice as the mean value of the saliency map Sal :

$$T_a = \frac{2}{W \times H} \sum_{i=1}^W \sum_{j=1}^H |Sal(i, j)|. \quad (26)$$

Results are shown in Figures 11(b), 12(b), 13(b), and 14(b). Our method achieves the highest precision and the second best F-measure on the two largest datasets MSRA10K and ECSSD and the most challenging dataset Judd further proves the effectiveness of our proposed method, for the precision is the ratio of the correctly detected areas to the whole detected areas. Fourth best precision and F-measure for our method are observed on SED2. For SED2 whose images contain two objects in more complex background, the performance of our

FIGURE 16: Qualitative comparisons of our saliency maps and recently developed approaches on ECSSD, Judd, and SED2 datasets.

method is close to GMR. Another major factor affecting the performance of the GMR algorithm is that SED2 has many images whose labeled objects touch boundaries leading to failure in satisfying the boundary prior used in GMR. In such cases, it may be better to carry out a rough detection using only contrast rather than taking advantage of the boundary prior knowledge. This reveals the reason that SVO performs inferiorly on other datasets, while it achieves relatively good performance on SED2.

MAE. Since precision-recall curves do not reflect the high-light level of an entire object, we also compute the MAE

values. As Figures 11(c), 12(c), 13(c), and 14(c) show, our method consistently produces the minor error on all four datasets. Specifically, it ranks the least error on ECSSD and the second least error on MSRA10K, Judd, and SED2, which indicates our method with a more robust performance on different datasets. In addition, despite a relatively good performance in precision-recall curves and F-measure, SVO has the largest MAE due to its weak background suppression.

Summary. Though our method does not exhibit the top best effect on all four datasets experiment, however, it with

TABLE 1: Execution time comparison.

Method (Matlab)	CA	CB	GMR	GR	HS	LMLC	PCA	SVO	Ours
Time (s)	40.9	2.24	0.149	1.35	0.528	140.0	4.34	56.5	6.14

a consistently excellent and robust performance compared with the other eight approaches.

5.3. Qualitative Comparisons. In addition to quantitative comparison, we also, respectively, provide the visual comparison of different methods in Figures 15 and 16, from which we can see that our method generates the best detection results even in some challenging cases such as the background of image being very extremely cluttered or the image with complex texture. For example, in rows 3, 7, and 10 of Figure 15, our approach can generate reasonable saliency map and more successfully highlights the entire salient object, while the other method may be distracted and influenced by the cluttered background textures.

Comparatively speaking, we can clearly observe the distinctness of different type of models, and a good detection approach tries to highlight the whole salient object while suppressing the background. Some models that incorporate a center-bias component also produce appealing results, for example, CB. However, this method suffers failure when the objects locate far off the image center. Interestingly, compared with other pixel-based or patch-based approaches, those region-based models, for example, CB, HS, GMR, GR, and ours, always preserve the object boundary well. PCA detects the outlines of the saliency objects while missing the interiors as it relies mostly on patterns. Though LMLC and SVO detect the salient regions, parts of the background are mistakenly detected as salient. HS is easy to miss a small target, such as the first row in Figure 16.

Another side-effect of some method is that it is possible to lose useful object parts if it involves thresholding operation. That happens occasionally on SED2 dataset (e.g., 11th row in Figure 16). One of the two objects in image can be missed after threshold, as two objects probably with different saliency levels. In contrast, such misjudgment can reduce in our method as no threshold is used to binarize the saliency map produced in intermediate procedure like in GMR.

5.4. Computation Cost Analysis. Table 1 shows the average execution time of processing one image over all 10K images of MSRA10K (typical 400×300 image resolution). Experiments are conducted on an Intel Xeon E5645 2.40 GHz machine equipped with 8 GB RAM. LMLC takes the longest time because of its Laplacian sparse subspace clustering process. SVO also spent a longer time for it introduces an objectness measure for each single image. The most time consuming step in our method is the process of coarse salient regions extraction, which takes about 3.57 s (51%).

6. Conclusions

In this work, we proposed a new approach first which comprehensively utilizes boundary prior information and local and global information to extract a variety of saliency features

from an image to construct a coarse saliency map, then obtain a background map by means of boundary labels propagation, and finally integrate these two maps into a unified optimization framework to acquire a refined saliency detection result. We validate and compare our method on four popular databases against 8 mainstream approaches. The experimental results demonstrate that our method exhibits refined and consistently excellent saliency detection.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

The paper is supported by the National Science and Technology Support Project under Grant 2015IM030300.

References

- [1] G. R. Mangun, "Neural mechanisms of visual selective attention," *Psychophysiology*, vol. 32, no. 1, pp. 4–18, 1995.
- [2] A. Borji, M.-M. Cheng, H. Jiang, and J. Li, "Salient object detection: a benchmark," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5706–5722, 2015.
- [3] Z. Ren, S. Gao, L.-T. Chia, and I. Tsang, "Region-based saliency detection and its application in object recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 5, pp. 769–779, 2014.
- [4] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 185–198, 2010.
- [5] H. Huang, L. Zhang, and H.-C. Zhang, "Arcimboldo-like collage using internet images," *ACM Transactions on Graphics*, vol. 30, no. 6, article 155, 2011.
- [6] A. Li, X. She, and Q. Sun, "Color image quality assessment combining saliency and fsm," in *Proceedings of the 5th International Conference on Digital Image Processing (ICDIP '13)*, vol. 8878, Beijing, China, April 2013.
- [7] C. Qin, G. Zhang, Y. Zhou, W. Tao, and Z. Cao, "Integration of the saliency-based seed extraction and random walks for image segmentation," *Neurocomputing*, vol. 129, pp. 378–391, 2014.
- [8] S. Feng, D. Xu, and X. Yang, "Attention-driven salient edge(s) and region(s) extraction with application to CBIR," *Signal Processing*, vol. 90, no. 1, pp. 1–15, 2010.
- [9] L. Marchesotti, C. Cifarelli, and G. Csurka, "A framework for visual saliency detection with applications to image thumbnailing," in *Proceedings of the 12th IEEE International Conference on Computer Vision (ICCV '09)*, pp. 2232–2239, IEEE, Kyoto, Japan, September 2009.
- [10] J. Yang and M.-H. Yang, "Top-down visual saliency via joint CRF and dictionary learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 2296–2303, IEEE, Providence, RI, USA, June 2012.

- [11] C. Kanan, M. H. Tong, L. Zhang, and G. W. Cottrell, "SUN: top-down saliency using natural statistics," *Visual Cognition*, vol. 17, no. 6-7, pp. 979-1003, 2009.
- [12] M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S. M. Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569-582, 2015.
- [13] M. M. Cheng, G. X. Zhang, N. J. Mitra, X. Huang, and S. M. Hu, "Global contrast based salient region detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 409-416, 2011.
- [14] E. Rahtu, J. Kannala, M. Salo et al., *Segmenting Salient Objects from Images and Videos*, Computer Vision-ECCV 2010, Springer, Berlin, Germany, 2010.
- [15] Y. Xie and H. Lu, "Visual saliency detection based on Bayesian model," in *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP '11)*, pp. 645-648, IEEE, Brussels, Belgium, September 2011.
- [16] N. Bruce and J. Tsotsos, "Saliency based on information maximization," in *Advances in Neural Information Processing Systems 18 (NIPS 2005)*, pp. 155-162, MIT Press, 2005.
- [17] T. Liu, Z. Yuan, J. Sun et al., "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353-367, 2011.
- [18] D. Gao, V. Mahadevan, and N. Vasconcelos, "The discriminant center-surround hypothesis for bottom-up saliency," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS '07)*, pp. 497-504, 2007.
- [19] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254-1259, 1998.
- [20] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in Neural Information Processing Systems*, pp. 545-552, 2006.
- [21] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk, "Salient region detection and segmentation," in *Computer Vision Systems*, A. Gasteratos, M. Vincze, and J. K. Tsotsos, Eds., vol. 5008 of *Lecture Notes in Computer Science*, pp. 66-75, Springer, Berlin, Germany, 2008.
- [22] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1915-1926, 2012.
- [23] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk, "Frequency-tuned salient region detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 1597-1604, IEEE, Miami, Fla, USA, June 2009.
- [24] X. Hou and L. Zhang, "Saliency detection: a spectral residual approach," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1-8, Minneapolis, Minn, USA, June 2007.
- [25] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569-582, 2015.
- [26] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Computer Vision—ECCV 2012*, vol. 7574, pp. 29-42, Springer, Berlin, Heidelberg, 2012.
- [27] Y. Fang, W. Lin, C. T. Lau, and B.-S. Lee, "A visual attention model combining top-down and bottom-up mechanisms for salient object detection," in *Proceedings of the 36th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '11)*, pp. 1293-1296, IEEE, Prague, Czech Republic, May 2011.
- [28] B. Rasolzadeh, A. T. Targhi, and J. O. Eklundh, "An attentional system combining top-down and bottom-up influences," in *Attention in Cognitive Systems. Theories and Systems from an Interdisciplinary Viewpoint*, pp. 123-140, Springer, Berlin, Germany, 2007.
- [29] H. Tian, Y. Fang, Y. Zhao, W. Lin, R. Ni, and Z. Zhu, "Salient region detection by fusing bottom-up and top-down features extracted from a single image," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4389-4398, 2014.
- [30] L. Xu, Q. Yan, Y. Xia, and J. Jia, "Structure extraction from texture via relative total variation," *ACM Transactions on Graphics*, vol. 31, no. 6, article 139, 2012.
- [31] F. W. Campbell and J. G. Robson, "Application of Fourier analysis to the visibility of gratings," *The Journal of Physiology*, vol. 197, no. 3, pp. 551-566, 1968.
- [32] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274-2282, 2012.
- [33] J. Chen, S. Shan, C. He et al., "WLD: a robust local image descriptor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1705-1720, 2010.
- [34] A. K. Jam, "Fundamentals of digital image processing," in *Information and System Sciences*, Prentice-Hall, 1989.
- [35] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America A*, vol. 4, no. 12, pp. 2379-2394, 1987.
- [36] L. Zhang, Z. Gu, and H. Li, "SDSP: a novel saliency detection method by combining simple priors," in *Proceedings of the 20th IEEE International Conference on Image Processing (ICIP '13)*, pp. 171-175, Melbourne, Australia, September 2013.
- [37] H. Li, H. Lu, Z. Lin, X. Shen, and B. Price, "Inner and inter label propagation: salient object detection in the wild," *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 3176-3186, 2015.
- [38] C. Zhao, C. C. Liu, Z. Lai, H. Rao, and Z. Li, "Sparse embedding visual attention system combined with edge information," *AEU—International Journal of Electronics and Communications*, vol. 65, no. 12, pp. 1061-1068, 2011.
- [39] L. Itti and C. Koch, "Feature combination strategies for saliency-based visual attention systems," *Journal of Electronic Imaging*, vol. 10, no. 1, pp. 161-169, 2001.
- [40] Q. Zhao and C. Koch, "Learning visual saliency by combining feature maps in a nonlinear manner using AdaBoost," *Journal of Vision*, vol. 12, no. 6, article 22, 2012.
- [41] T. Judd, K. Ehinger, F. Durand et al., "Learning to predict where humans look," in *Proceedings of the IEEE 12th International Conference on Computer Vision (ICCV '09)*, pp. 2106-2113, IEEE, Kyoto, Japan, 2009.
- [42] S. Alpert, M. Galun, A. Brandt, and R. Basri, "Image segmentation by probabilistic bottom-up aggregation and cue integration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 315-327, 2012.
- [43] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 1155-1162, Portland, Ore, USA, June 2013.
- [44] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proceedings*

of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13), pp. 3166–3173, IEEE, Portland, Ore, USA, June 2013.

- [45] Y. Xie, H. Lu, and M.-H. Yang, “Bayesian saliency via low and mid level cues,” *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1689–1698, 2013.
- [46] K.-Y. Chang, T.-L. Liu, H.-T. Chen, and S.-H. Lai, “Fusing generic objectness and visual saliency for salient object detection,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 914–921, Barcelona, Spain, November 2011.
- [47] R. Margolin, A. Tal, and L. Zelnik-Manor, “What makes a patch distinct?” in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 1139–1146, IEEE, Portland, Ore, USA, June 2013.
- [48] H. Jiang, J. Wang, Z. Yuan, T. Liu, and N. Zheng, “Automatic salient object segmentation based on context and shape prior,” in *Proceedings of the British Machine Vision Conference (BMVC '11)*, pp. 110.1–110.12, BMVA Press, September 2011.
- [49] C. Yang, L. Zhang, and H. Lu, “Graph-regularized saliency detection with convex-hull-based center prior,” *IEEE Signal Processing Letters*, vol. 20, no. 7, pp. 637–640, 2013.

Hindawi

Submit your manuscripts at
<http://www.hindawi.com>