

A Hybrid of Local and Global Saliencies for Detecting Image Salient Region and Appearance

Qinmu Peng, Yiu-ming Cheung, *Senior Member, IEEE*, Xinge You, *Senior Member, IEEE*,
and Yuan Yan Tang, *Fellow, IEEE*

Abstract—This paper presents a visual saliency detection approach, which is a hybrid of local feature-based saliency and global feature-based saliency (simply called local saliency and global saliency, respectively, for short). First, we propose an automatic selection of smoothing parameter scheme to make the foreground and background of an input image more homogeneous. Then, we partition the smoothed image into a set of regions and compute the local saliency by measuring the color and texture dissimilarity in the smoothed regions and the original regions, respectively. Furthermore, we utilize the global color distribution model embedded with color coherence, together with the multiple edge saliency, to yield the global saliency. Finally, we combine the local and global saliencies, and utilize the composition information to obtain the final saliency. Experimental results show the efficacy of the proposed method, featuring: 1) the enhanced accuracy of detecting visual salient region and appearance in comparison with the existing counterparts, 2) the robustness against the noise and the low-resolution problem of images, and 3) its applicability to multisaliency detection task.

Index Terms—Gradient minimization, multiple salient edges, saliency detection, visual attention.

Manuscript received November 20, 2015; revised January 15, 2016 and March 16, 2016; accepted April 18, 2016. Date of publication May 26, 2016; date of current version December 14, 2016. This work was supported in part by the Faculty Research Grant of Hong Kong Baptist University (HKBU) under Project FRG1/14-15/041, Project FRG2/14-15/075, and Project FRG2/15-16/049, in part by the HKBU Knowledge Transfer Office under Grant MPCF-005-2014/2015, in part by the National Natural Science Foundation of China under Grant 61272366, Grant 61272203, and Grant 61273244, in part by the National Technologies R&D Program under Grant 2015BAK36B00, in part by the Hubei Province Technologies Research and Development Program under Grant XYJ2014000459, in part by the Research Grants of University of Macau under Grant MYRG2015-00049-FST, Grant MYRG2015-00050-FST, and Grant RDG009/FST-TYY/2012, in part by the Science and Technology Development Fund of Macau under Grant 100-2012-A3 and Grant 026-2013-A, and in part by the Macau-China Join Project under Grant 008-2014-AMJ. This paper was recommended by Associate Editor M. Celenk. (*Corresponding author: Yiu-ming Cheung.*)

Q. Peng is with the Department of Computer Science, Hong Kong Baptist University, Hong Kong (e-mail: pqinmu@gmail.com).

Y.-m. Cheung is with the Department of Computer Science, Hong Kong Baptist University, Hong Kong, and also with United International College, Beijing Normal University-Hong Kong Baptist University, Zhuhai 519000, China (e-mail: ymc@comp.hkbu.edu.hk).

X. You is with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: youxg@hust.edu.cn).

Y. Y. Tang is with the Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Macau 999078, China (e-mail: yytang@umac.mo).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSMC.2016.2564922

I. INTRODUCTION

VISUAL attention [1], [2] is an important characteristic in the human visual system (HVS), through which a person can effortlessly identify the interesting parts of a scene, i.e., visual saliency. This selective attention mechanism can facilitate the high-level cognitive task. Thus far, visual saliency has applied to a variety of application areas, e.g., object recognition [3], image/video compression [4], [5], image cropping/retargeting [6], [7], human gaze and actions [8], [9], image retrieval [10], and so forth.

With the goal to achieve a comparable saliency detection performance of HVS, a number of saliency models have been proposed in the past decades. Basically, most of these works seek to employ different features, which include the local and global features, to estimate the image saliency. The local feature-based approaches extract the visual features from pixels or patches in an image, such as [11]–[18]. For instance, Li *et al.* [15] extracted local visual attributes to characterize an image patch and proposed multi-instance learning for visual saliency estimation. By contrast, the global feature-based approaches, (see [19]–[28]), estimate the visual saliency from the whole image. For example, Guo and Zhang [24] utilized the phase spectrum of quaternion Fourier transform to generate the saliency predictions.

The saliency detection performance has been improved with the emerging saliency models. Nevertheless, saliency models based on either local features or global features alone are still insufficient to effectively handle the images with low contrast between object and background, complicated background, and noise.

With the main motivation to improve the overall saliency detection performance, we focus on seeking effective local and global features and combination model. Thus, there are two key issues to be addressed: one is how to choose and extract the local and global features, which should complement each other. The other issue is to design an optimal combination scheme for combining the different saliencies. The works in [29] and [30] utilize the equal weights only, which is, however, too simple to achieve more accurate visual saliency. Hence, it is still desirable to design the weights that can be adjustable for each individual image. To this end, this paper will address the above two issues so that the detection accuracy of visual saliency can be enhanced in terms of the salient region and appearance. For the first issue, our preliminary

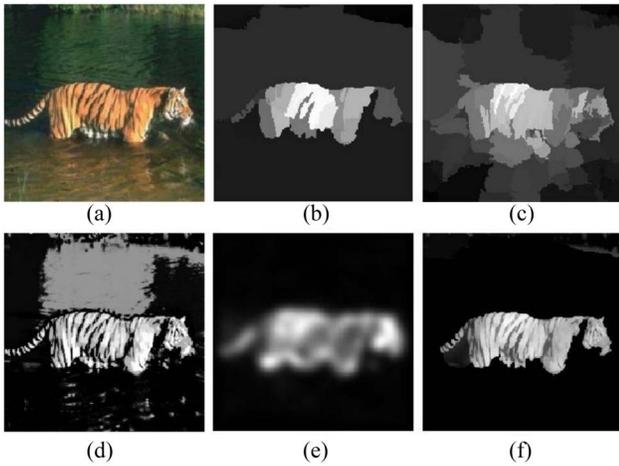


Fig. 1. Combination of the four saliency maps, in which: (a) input image, (b) local color saliency, (c) local textural saliency, (d) global color saliency, (e) global edge saliency, and (f) final saliency.

studies have found that the color, texture, edges or shapes are the important visual cues for image saliency detection. Accordingly, we shall choose and extract the corresponding features for saliency detection. Consequently, we will have two kinds of features: 1) local features and 2) global features, based on the different perspectives and the extraction methods. Specifically, the local features consist of local region-based color feature, and local region-based texture feature where the orientation information can be found as well, while the global features comprise: 1) global color feature that describes the color distribution in the whole image, and 2) global edge feature that is extracted in the frequency domain after multiscale smoothing, which can suppress the repeated patterns and preserve the edges. Hereinafter, the saliencies obtained using local region-based color feature, local region-based texture feature, global color feature, and global edge feature individually are named as local color saliency, local textural saliency, global color saliency, and global edge saliency, respectively. Further, the first two are simply called local saliency, whereas the latter two are called global saliency.

For the other issue, we will combine different saliency maps stated above to obtain the final saliency as illustrated in Fig. 1 via utilizing the conditional random field (CRF) learning, which can estimate the linear weights to get an optimal linear combination under the maximum likelihood criterion. Subsequently, considering the diversity of individual images, we further adopt a tuning scheme to adjust the weights to adapt to the variance of saliency compactness, and obtain the final saliency. In summary, the contribution of this paper is mainly twofold:

- 1) A smoothing scheme for the local color saliency detection and multiple edge saliency is proposed. Compared with the existing models, the proposed method can provide better saliency, and can work well even in the low-resolution or noisy images.
- 2) The combination of the local and global saliency scheme which employs different visual cues can make fully use of their individual merits to achieve improved performance.

II. RELATED WORK

Saliency estimation methods can be explored from the different perspectives. In general, the global features intend to capture the natural characteristics of an image, which can help describe the image as a whole. By contrast, the local features are to extract the contrast between the center pixel (or patch) and its neighbors. The strength of the contrast contributes to the saliency, but its validity is not guaranteed due to its local perspective. Evidently, global features are helpful to eliminate the false positives caused by adopting local features alone. That is, the merit of global features is complementary to that of the local features. Under the circumstances, it is expected that the utilization of the combined local and global features can enhance the detection accuracy of the visual saliency.

In this paper, we focus on a combination of local and global features. Accordingly, several works have been made in the literature. For example, the works in [29] and [31] follow the feature integration theory [32], and combine the saliency maps from low-level features (e.g., color, contrast, and orientation) on multiple scales to estimate the saliency map. Moreover, Goferman *et al.* [30] utilized the local and global factors and associated them with the high-level visual patterns (e.g., face) to obtain the context-aware saliency. Li *et al.* [33] utilized independent component analysis to learn local saliency from image patches, which are combined with the frequency-based global saliency to generate the final saliency. In addition, Liu *et al.* [34] utilized the multiscale contrast, center-surround histogram and color spatial distribution to learn a CRF for binary saliency estimation. Inspired by this paper, we adopt the CRF and further adopt a tuning scheme that adjusts the weights to improve the performance of saliency detection.

Recently, Liu *et al.* [35] have proposed a framework termed as saliency tree (ST). They generated the saliency map and defined the merging criterion and merging order. Based on the analysis of the tree structure, a final pixel-based saliency map was derived. Furthermore, Vilaplana [36] has proposed two saliency models for salient object segmentation based on a hierarchical image segmentation. The two models analyze saliency at the different levels of details in the hierarchy. Similarly, the proposed method also adopts the concept of multiple-scale to obtain the edge saliency map, which is capable of detecting the salient objects in different sizes, with accurate boundaries.

III. PROPOSED METHOD

The proposed approach consists of two phases as shown in Fig. 2. In Phase 1, given an input image, a smoothing procedure is utilized as a preprocessing for extracting the local color saliency and the global edge saliency, by which the image regions become more homogeneous and the edges can be preserved to benefit the saliency detection. Then, we will represent and extract the local and global features stated in Section I. Subsequently, the corresponding saliencies are obtained. Finally, Phase 2 will combine different saliency maps and output the final saliency. In the following sections, we will describe the details of each phase, respectively.

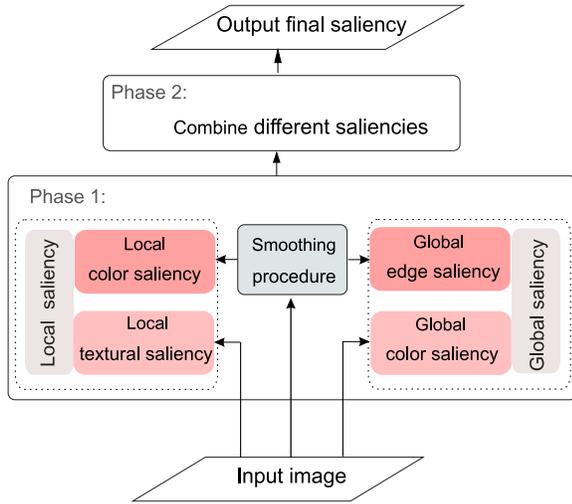


Fig. 2. Sketch of the implementation procedure of the proposed method.

A. Image Preprocessing: Smoothing Scheme With Automatic Selection of Smoothing Parameter

The work in [37] is a well-principled and efficient smoothing method based on the mechanism of counting spatial changes, which can obtain homogeneous regions and globally-preserved salient edges. Therefore, the foreground is emphasized and the background is more homogeneous after smoothing such as [38], which is of benefit to the saliency detection. In [37], the smooth parameter should be preassigned appropriately. Otherwise, the smoothing result becomes too blurry or remains unchanged when compared to the original image. However, this is not a trivial task from the practical viewpoint. Accordingly, we will propose a smoothing scheme based on [37], which can automatically select the smoothing parameter to filter out noise and repeated patterns, as well as keep the edges in the image. We utilize the L_0 gradient minimization smoothing

$$\min_{\mathcal{I}} \left\{ \sum_{(x,y)} (\mathcal{I}(x,y) - I(x,y))^2 + \lambda \cdot C(\mathcal{I}) \right\} \quad (1)$$

where $\mathcal{I}(x,y)$ denotes the value of the smoothed image \mathcal{I} at coordinates (x,y) , while $I(x,y)$ is the corresponding original value in the input image I . Furthermore, $C(\mathcal{I})$ counts the number of pixels, whose gradient magnitude is not zero, in \mathcal{I} . The smoothing parameter λ is a weight directly controlling the significance of $C(\mathcal{I})$.

In this paper, we propose an automatic method to calculate λ based on the sharpness of the image. Specifically, we measure the sharpness by utilizing the reconstruction error between the restored image and the original one. First, we convert an RGB image to YUV color space. The Y channel image (brightness-sensitive for the human eye) is then split into the blocks of 8×8 pixels. Subsequently, the discrete cosine transform (DCT) is performed. In generally, a person is capable of perceiving a small change in brightness rather than the exact strength of a high frequency brightness variation. Under the circumstances, quantization matrix can be used to take fewer bits to represent the information. A typical quantization matrix

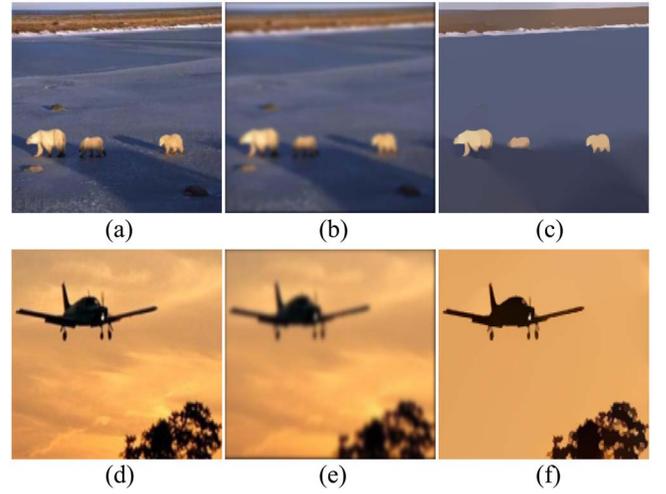


Fig. 3. Example of smoothing results, in which subfigure (a) and (d) are original images, (b) and (e) are the corresponding results via Gaussian smoothing, (c) and (f) are the smoothing results by the proposed method. Note that the original images contain small and big salient regions, respectively.

(for a quality of 50% as specified in the original JPEG standard) is utilized. We therefore restore the image by the inverse DCT transformation. Evidently, there exists the reconstruction error between the restored image and original image. The sharpness of an image can be indicated by the reconstruction error, i.e., larger reconstruction error implies more sharpness in the image. Subsequently, the average reconstruction error can be given as follows:

$$\epsilon = \min(9, \epsilon') \quad (2)$$

with

$$\epsilon' = \frac{\sum_{x,y} |I(x,y) - \mathbb{G}(x,y)|}{M \cdot N} \quad (3)$$

where $\mathbb{G}(x,y)$ is the value of the restored image \mathbb{G} at coordinates (x,y) , M and N are the width and height of an image, respectively, ϵ measures the sharpness of an image, and the maximum value of ϵ is set at 9 to avoid smoothing an image too much. Furthermore, the smooth parameter λ_{opt} is defined as $\lambda_{\text{opt}} = 0.01 \cdot \epsilon$ by a rule of thumb. An example of smoothed images is illustrated in Fig. 3. It can be seen that the image smoothed by the proposed method generates more homogeneous background than the one smoothed by Gaussian filter. Obviously, it will be of benefit to the saliency in images.

B. Detection of Local Saliency

In this section, we will elaborate how to obtain the local saliencies by utilizing the color and texture information contained in small regions, and how the saliency of one region is estimated based on its contrast to the other regions.

1) *Local Color Saliency*: Suppose we smooth the input image using the proposed method in Section III-A. Then, it is partitioned into regions using the mean shift algorithm [39]. The regions are denoted as $\{r_1, r_2, \dots, r_K\}$ and K is the number of total regions. We consider two factors to evaluate the local saliency: 1) the dissimilarity between image regions, and

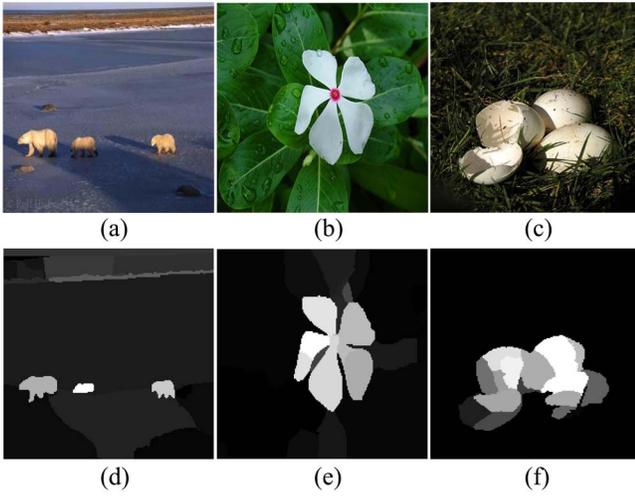


Fig. 4. Image saliency using the proposed local color saliency, in which subfigures (a)–(c) are input images and (d)–(f) are the corresponding local color saliencies, respectively.

2) the spatial distance. Specifically, for a region r_k , we calculate its saliency by measuring its color contrast to all other regions in LAB color space. $D(r_k, r_i)$ denoting the difference between region r_k and r_i is defined as follows:

$$D(r_k, r_i) = R(r_i)|\bar{u}_k - \bar{u}_i| \quad (4)$$

where $R(r_i)$ is the ratio between the area of region r_i and the whole image, \bar{u}_k and \bar{u}_i are the mean color in region r_k and r_i , respectively. By incorporating the element of spatial distance, i.e., Euclidean distance between the regions, the local color saliency of region r_k is given as

$$S_{lc}(r_k) = v_1 \cdot \sum_{i \neq k} v_2 \cdot D(r_k, r_i) \quad (5)$$

$$v_1 = \exp(-\sigma_1 \cdot \text{Dist}(r_k, I_c)) \quad (6)$$

$$v_2 = \exp(-\sigma_2 \cdot \text{Dist}(r_k, r_i)) \quad (7)$$

where v_1 and v_2 are two different weights denoting the distance affects. I_c is the image center, and $\text{Dist}(r_k, I_c)$ is the spatial distance between r_k and I_c , $v_2 = \exp(-\sigma_2 \cdot \text{Dist}(r_k, r_i))$, and $\sigma_1 = \sigma_2 = 0.005$. As a result, the saliency of the image is obtained by assigning $S_{lc}(r_k)$ to the corresponding region r_k , with $k = 1, 2, \dots, K$. A snapshot of the image saliency result using the proposed local color saliency is illustrated in Fig. 4. It can be seen that visual saliency has been detected successfully.

2) *Local Textural Saliency*: We need to take into account the texture information which is ubiquitous in natural images and constitutes an important visual cue to the image saliency. In the literature, there are a number of works on the texture features; e.g., multifractal texture [40], generic texture [41], local binary pattern (LBP) [42], [43], and so on. Most of them have achieved great success in their application domains. According to the theory of HVS, visual attention is closely related to the visual cues [44] such as arch, corner, and circle, which have specific geometry shapes that capture more attention. Among

example			threshold			weights		
14	13	12	1	0	0	1	2	4
15	14	11	1		0	128		8
16	17	18	1	1	1	64	32	16

$$\begin{aligned} \text{Patten} &= 10001111 \\ \text{LBP} &= 1+16+32+64+128 = 241 \end{aligned}$$

Fig. 5. Example of calculating the LBP with a block of 3×3 pixels, and the enumeration starts at the top-left corner. The gray value of the center pixel is 14 and its LBP value is 241.

the above texture features, LBP is a robust and efficient texture feature descriptor which can effectively extract the arch, corner and circle texture feature for the saliency detection task.

In this paper [42], there are 256 LBPs to represent textures and LBP is defined as follows:

$$\text{LBP} = \sum_{n=0}^7 2^n \cdot \phi(i_n - i_c) \quad (8)$$

where n runs over the eight neighbors of the central pixel c , i_n and i_c are the gray value, and $\phi(\mu)$ is 1 if $\mu \geq 0$ and 0 otherwise. An example of calculating LBP is shown in Fig. 5. Among the 256 patterns, 58 of them provide a vast majority of local texture patterns that are called “uniform” patterns, which means that there are a limited number of transitions in the circular presentation of the pattern [45]. As for the saliency detection task, we should seek the visual patterns that contribute the most to the visual saliency. To this end, 18 patterns are selected from the uniform patterns. We call them selective LBPs (SLBPs) as shown in Fig. 6, which are more useful than the other patterns to represent the specific geometry shapes, i.e., arch, corner and circle. Meanwhile, these selective patterns with a smaller size can save the computational cost moderately in the saliency detection, which is an additional benefit.

In the previous section, the image is partitioned into the different regions for the local color saliency detection. The texture dissimilarity is also computed on the regions. Accordingly, the LBP is calculated for pixels in the image (excluding the border pixels). Then, we compute the histograms for the SLBP within each region, i.e., counting the number of each type of SLBP within a region. Subsequently, χ^2 distance is adopted to compute the distance between two histograms

$$\chi^2(p, q) = \sum_i \frac{(p_i - q_i)^2}{p_i + q_i} \quad (9)$$

where p and q are the two different histograms, p_i and q_i are the i th element in p and q , respectively.

For each region r_k , we calculate its saliency by measuring its texture contrast to all other regions in LAB color space. $\mathbb{D}(r_k, r_i)$ denoting the difference between region r_k and r_i is defined as follows:

$$\mathbb{D}(r_k, r_i) = \chi^2(\bar{p}^{r_k}, \bar{q}^{r_i}) \quad (10)$$

where \bar{p}^{r_k} and \bar{q}^{r_i} are the SLBP histograms in region r_k and r_i , respectively. Considering the spatial distance, the local textural

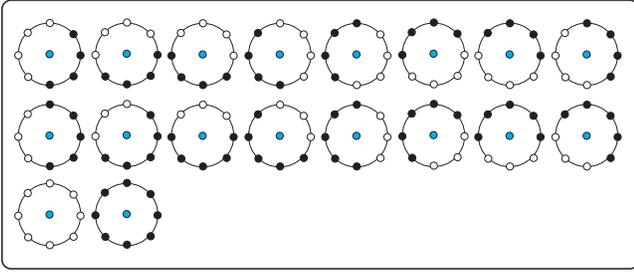


Fig. 6. Selected LBPs, which contain corner feature, arch feature, and the circle feature.

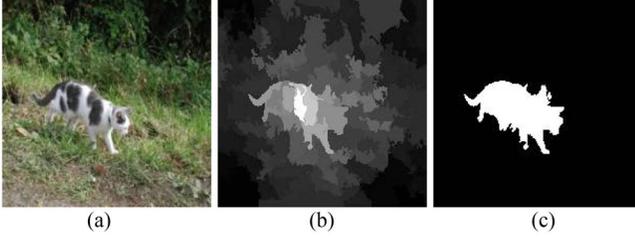


Fig. 7. Image saliency using the proposed local textural saliency, where (a) input image, (b) local textural saliency, and (c) binary image of local textural saliency using the threshold.

saliency of region r_k is given as

$$S_{lt}(r_k) = v_1 \cdot \sum_{i \neq k} v_2 \cdot \mathbb{D}(r_k, r_i) \quad (11)$$

where v_1 and v_2 are the same as the ones in (5), and $\sigma_1 = \sigma_2 = 0.005$. As a result, the saliency of the image is obtained by assigning $S_{lt}(r_k)$ to the corresponding region r_k , with $k = 1, 2, \dots, K$. An example based on the local saliency is shown in Fig. 7, in which the input image includes various textures such as *Grass*, *Leaf*, *Earth*, and a *Cat's* fur. The binary image obtained by thresholding the textural saliency map presents the shape of a *Cat*. It can be seen that salient regions have been detected successfully.

C. Detection of Global Saliency

To capture the saliency of an image as a whole, we perform the detection of global saliency. Specifically, the global color spatial-distribution is embedded with color coherence, and the multiscale edge density is utilized as the visual cue for the saliency detection task.

1) *Global Color Saliency*: In [34], it utilizes the color spatial distribution to describe the saliency of an object upon the fact that the wider a color spreads in the image, the less possible a salient object contains this color. In this paper, based on the color spatial distribution, we further take into account the color coherence that measures the color similarity in small blocks, which provides a better description for the color saliency.

Specifically, we fit a Gaussian mixture model

$$p(I_x) = \sum_{c=1}^C \omega_c \cdot N(I_x | \mu_c, \Sigma_c) \quad (12)$$

for the colors in the image, where ω_c is the weight, and $N(I_x | \mu_c, \Sigma_c)$ is the c th Gaussian component with the mean μ_c and the covariance matrix Σ_c . Also, C is the number of the color components. In this paper, we simply set C at 6. Each pixel x represented by RGB color features is assigned to a color component with the probability

$$p(c|I_x) = \frac{\omega_c \cdot N(I_x | \mu_c, \Sigma_c)}{\sum_c \omega_c \cdot N(I_x | \mu_c, \Sigma_c)}. \quad (13)$$

To evaluate the saliency of each color component, the variance of color spatial-distribution and color coherence are computed to measure the saliency. That is, the horizontal variance is given as

$$V_h(c) = \frac{1}{\sum_x p(c|I_x)} \sum_x p(c|I_x) \cdot \|x_h - M_h(c)\|^2 \quad (14)$$

$$M_h(c) = \frac{1}{\sum_x p(c|I_x)} \sum_x p(c|I_x) \cdot x_h \quad (15)$$

where x_h is the x -coordinate of the pixel x . The spatial variance of the component c is

$$V(c) = V_h(c) + V_v(c) \quad (16)$$

where $V_h(c)$ is the horizontal variance and $V_v(c)$ is the vertical variance, which can be defined similarly to $V_h(c)$. Then, we normalize $V(c)$ to $[0, 1]$, i.e., $V(c) \leftarrow (V(c) - \min(V(c))) / (\max(V(c)) - \min(V(c)))$. As for the color coherence, it is found that the color coherence is small when two color components are mixed. That is, a pixel with larger color coherence has a higher probability of being saliency. Hence, denoting $H(c)$ as the color distribution difference, we define the color coherence as $1 - H(c)$, and $H(c)$ is calculated by the following expression:

$$H(c) = \frac{1}{\sum_x p(c|I_x)} \sum_x p(c|I_x) \nabla(P_{N_x}) \quad (17)$$

where N_x denotes the 8-point neighborhood set of the pixel x , and $\nabla(P_{N_x}) = \max(p(c|I_x) | x \in N_x) - \min(p(c|I_x) | x \in N_x)$ denotes the difference between the maximum and minimum value of $p(c|I_x)$ in N_x . Then, $H(c)$ is normalized to $[0, 1]$ analogous to $V(c)$. Accordingly, the global color saliency S_{gc} is defined as

$$S_{gc} = \sum_c (1 - V(c))(1 - H(c)) \mathcal{D}(c) p(c|I_x) \quad (18)$$

where $\mathcal{D}(c) = \sum_x p(c|I_x) \exp(-\tau d_x)$ is a weight which assigns less importance to the color far from the image center. τ is set at a small value, say $\tau = 0.005$, and d_x is the distance from the pixel x to the image center. $\mathcal{D}(c)$ is also normalized to $[0, 1]$ similarly to $V(c)$. To show the effectiveness of S_{gc} , Fig. 8 illustrates an example. It can be seen that, compared to the global color saliency of [34], the proposed S_{gc} achieves more desirable saliency for the object in the image when considering color coherence. It represents the good property of increasing the saliency of pixels that share the color similarity with the nearby pixels. Therefore, the proposed global color saliency can achieve more homogeneous color saliency, which contributes much to the final saliency.

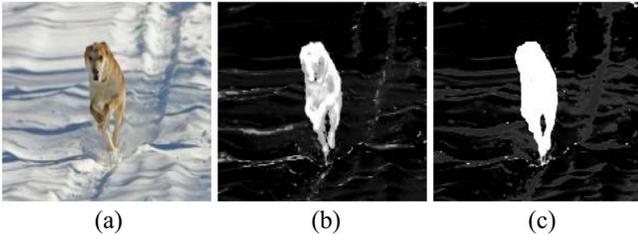


Fig. 8. Image saliency using the proposed global color saliency, where (a) input image, (b) global color saliency of [34], and (c) proposed global color saliency.

2) *Global Edge Saliency*: Edge density is another important visual cue for the image saliency, which has been utilized in the object detection and other aspects. There exist some popular edge operators such as Canny and Sobel operators, but they cannot be applied directly to natural images for the saliency detection task because the complicated texture and noisy pixels within images often result in the undesirable edges. To tackle this problem, the basic idea that we adopt in this paper is to filter out the unimportant or repeated texture in images, meanwhile preserving clear edges within the salient regions or objects. Actually, in the previous section, the proposed smoothing procedure has demonstrated the desirable property in dealing with this problem.

Accordingly, we implement a multilevel smoothing procedure on the image and extract visual edges in the frequency domain. This procedure incorporates the edge cues from the different smooth levels of the image to generate salient edge maps. It is expected that the combination of different edge maps from multiple images at different smoothing levels can alleviate the impact of false saliency caused by noise, and highlight the true saliency existing at each level. Specifically, suppose the images obtained at m different levels of smoothing are denoted as $\mathcal{I}_1, \dots, \mathcal{I}_m$, where \mathcal{I}_m is the smoothest image. In general, these images can be obtained using the different smoothing parameter $\lambda_k = k\lambda_{\text{opt}}$, with $k = 1, 2, \dots, m$, where λ_{opt} is given in Section III-A. In this paper, the level m is a small number and is set at 3. That is, we utilize progressively stronger parameters λ 's to generate images at different smoothing levels. In order to capture more robust edge maps, we perform the analysis of amplitude and phase spectrum on each smooth level of image. Differing from the work in [21] and [33], the proposed approach using multiple images at different smoothing levels can make full use of image information at different smoothing levels, which performs better than the spectrum residual [21], [33]. The edge saliency map of image \mathcal{I}_i is denoted as $S(\mathcal{I}_i)$ with

$$S(\mathcal{I}_i) = \left| F^{-1} \left\{ A(u, v) \cdot e^{iP(u, v)} \right\} \right|^2 \quad (19)$$

where $A(u, v) = |F\{\mathcal{I}_i(x, y)\}| * g_1$ is the smoothed amplitude spectrum of \mathcal{I}_i using a Gaussian filter $g_1 = \exp(-(u^2 + v^2)/\sigma^2)$, with $\sigma^2 = 0.05$, which suppresses nonedge patterns. $P(u, v)$ is the phase spectrum of the original image, $F\{\cdot\}$ is the Fourier transform, and $F^{-1}\{\cdot\}$ is the inverse Fourier transform.

To combine the different edge maps, we adopt the harmonic mean [20] because it has the property of increasing

Algorithm 1 Global Edge Saliency

Input: An image and the multiple levels m 's.

- 1: Smooth the image using the different parameters λ 's;
- 2: Obtain the multiple edge saliency maps using Eq.(19);
- 3: Merge the edge saliency maps by Eq.(20), and smooth it using Gaussian filter.

Output: Global edge saliency S_{ge} .

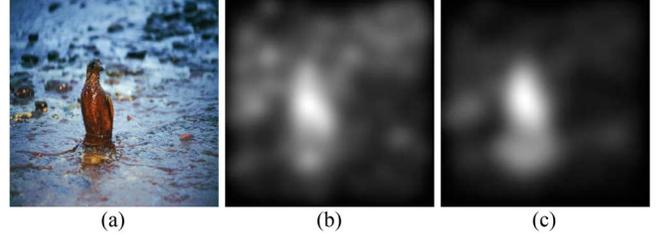


Fig. 9. Global edge saliency, where (a) input image, (b) saliency obtained by the spectrum scale-space analysis [33], and (c) saliency obtained by the proposed global edge saliency method.

the saliency of pixels when they are salient at multiple levels. The saliency edges obtained from the different smoothing parameters are combined by

$$S_{\text{multi}} = \left\{ \frac{1}{m} \sum_{i=1}^m \frac{1}{S(\mathcal{I}_i)} \right\}^{-1} \quad (20)$$

where S_{multi} is the merged saliency edge, which provides more robust edge saliency using the multiple smoothing schemes. After the combined saliency edge is obtained, the Gaussian filter g_2 with zero mean and variance $\sigma^2 = 0.05$ is utilized to blur the saliency edge map to generate the visual edge saliency S_{ge} , i.e., $S_{ge} = S_{\text{multi}} * g_2$. Algorithm 1 summarizes the procedure of acquiring the global edge saliency.

An example of the proposed global edge saliency compared with the spectrum scale-space analysis method [33] is illustrated in Fig. 9. It can be seen that the saliency generated by the proposed method is more desirable and the false saliency is largely suppressed.

D. Combination of Different Saliencies

We denote the previously stated four local and global saliencies, i.e., local color saliency, local textural saliency, global color saliency, and global edge saliency, as S_1, S_2, S_3 , and S_4 , respectively. These visual saliencies are complementary to each other, thus an integration of them appears promising. Although some works adopt equal weights to combine different saliency maps for simplicity, it is expected that tuning the weights can find a higher correlation between these visual saliencies, whereby improving the detection accuracy of saliency.

We train the CRF model to obtain the optimal combination weights w_i for S_i . Generally, once the weights are learned through CRF model, they are applied to all the new images without considering the performance variation on individual images. To improve the performance of saliency detection, we further adopt a tuning scheme that adjusts the weights to be

customized to individual image. Specifically, if the saliency map S_i , $i = 1, 2, 3, 4$ is more compact, it should be assigned a larger weight. Here, the spatial variance V_i measures the compactness α_i of each saliency map S_i

$$V_i = V_{ih} + V_{iv} \quad (21)$$

$$V_{ih} = \frac{\sum_x S_{ix} \cdot (x_{ih} - u_{ih})^2}{\sum_x S_{ix}} \quad (22)$$

$$u_{ih} = \frac{\sum_x S_{ix} \cdot x_{ih}}{\sum_x S_{ix}} \quad (23)$$

where V_{ih} is the horizontal variance, and V_{iv} is the vertical variance. S_{ix} and x_{ih} are the saliency value and the x -coordinate of the pixel x in S_i , respectively. $\alpha_i = (V_i / \sum_i V_i)$ denotes the compactness of each saliency map. Hence, the adjusted weight parameters are given as

$$\tilde{w}_i = \frac{w_i \cdot (1 - \alpha_i)}{\sum_i w_i \cdot (1 - \alpha_i)}. \quad (24)$$

Subsequently, the combination of the saliency map S_{all} is expressed as

$$S_{\text{all}} = \sum_{i=1}^4 \tilde{w}_i \cdot S_i. \quad (25)$$

After the merge of the different visual saliencies, we can implement the postprocessing to refine the saliency result by utilizing the composition information of the image regions. According to the HVS, it is found that a person generally pays less attention to the areas close to the boundary. In fact, most salient regions are located near the image center, and uninteresting regions like background often spread to the image boundary. Accordingly, we reduce the saliency of a region using an exponential function. If a region reaches any two horizontal or vertical image borders at the same time (denoted as condition 1 hereinafter), we reduce the region saliency by $e^{-A_k/\sigma_3 A_I}$, where A_k and A_I are the area of the k th region and whole image, respectively. Otherwise, we just reduce the region saliency by $e^{-n_k/n\sigma_4}$, i.e., the ratio between the joint border pixels n_k of the region and the all border pixels n of an image. In our implementation, we set $\sigma_3 = \sigma_4 = 0.1$. Therefore, $F_{ci}(\tilde{r}_k)$ denoting the weight for the region \tilde{r}_k is defined as follows:

$$F_{ci}(\tilde{r}_k) = \begin{cases} e^{-A_k/0.1A_I} & \text{if condition 1 holds} \\ e^{-n_k/0.1n} & \text{otherwise.} \end{cases} \quad (26)$$

As a result, the weight provided by the composition information for the image is obtained by assigning $F_{ci}(\tilde{r}_k)$ to the corresponding region \tilde{r}_k , with $k = 1, 2, \dots, K$. Finally, we obtain the overall saliency S_{final} by

$$S_{\text{final}} = S_{\text{all}} \cdot F_{ci}. \quad (27)$$

To demonstrate the implementation for the saliency detection, we summarize the whole procedure in Algorithm 2, in which steps 1–4 estimate the saliency based on the local and global features, respectively. Step 5 combines the different saliency through CRF and adjusts the weights to obtain the overall saliency. The last step, i.e., step 6, is a postprocessing procedure to refine the final saliency.

Algorithm 2 Image Saliency Detection

Input: Given an image I

- 1: Calculate the local color saliency using Eq.(5);
- 2: Calculate the local textural saliency using Eq.(11);
- 3: Calculate the global color saliency using Eq.(18);
- 4: Calculate the global edge saliency using *Algorithm 1*;
- 5: Combination of the above four saliency, i.e.,
 $S_{\text{all}} = \sum_{i=1}^4 \tilde{w}_i \cdot S_i$;
- 6: Refine the final saliency through Eq.(27), i.e.,
 $S_{\text{final}} = S_{\text{all}} \cdot F_{ci}$.

Output: Visual saliency S_{final} .

IV. EXPERIMENTS

A. Experimental Setting

1) *Data Sets:* In this section, we evaluate the performance of the proposed method on two public image datasets. The first dataset is MSRA10K [46], which is composed of 10000 images labeled with the pixelwise ground truth. The second dataset is extended complex scene saliency dataset (ECSSD) [17] consisting of 1000 images with the diverse patterns in both foreground and background, which is also labeled with the pixelwise ground truth.

2) *Baselines:* In the experiments, the proposed approach, denoted as OUR, is performed on the MSRA10K and ECSSD datasets in comparison with seven state-of-the-art algorithms, i.e., frequency-tuned model (FT) [22], saliency filters (SF) [47], PCA-based saliency (PCAS) [28], hierarchical saliency (HS) [17], saliency-based discriminative regional feature integration (DRFI) [18], ST [35], and saliency based on hierarchies (SOHs) [36].

3) *Evaluation Metrics:* To measure the performance, we will utilize not only the Precision, Recall, and F -measure as evaluation criteria to compare the performance of the above-stated algorithms, but also the mean absolute error (MAE) as another evaluation criterion, which was proposed by Perazzi *et al.* [47]

$$\text{MAE} = \frac{1}{|I|} \sum_x |S(x) - G(x)| \quad (28)$$

where S is the saliency map, G is the ground truth, and $|I|$ is the number of pixels (x) in the image I .

B. Experimental Results

First, we investigate the value selection of the smoothing parameter λ_{opt} for the smoothing scheme. In this paper, the local color saliency is calculated based on a subset (i.e., 1000 images) of MSRA10K dataset, and we find that the *precision* of saliency detection is decreased when the image is over-smoothed. We can utilize the *precision* of saliency detection (\hat{p}_1) obtained from the local color feature to find the suitable smoothing parameter for the input image (see Table I). Therefore, by a rule of thumb, we set $\lambda_{\text{opt}} = 0.01 \cdot \epsilon$ in our experiments, where ϵ is calculated by (2) according to the sharpness of an image in Section III-A.

Then, to evaluate the performance, we plotted the Precision–Recall (P–R) curves of the different methods. Further, MAE and F -measure were computed to evaluate these methods.

TABLE I
SMOOTHING PARAMETER, WHERE \hat{p}_1 IS THE
PRECISION OF SALIENCY DETECTION

$\frac{\lambda_{opt}}{\epsilon}$	0.002	0.006	0.01	0.014	0.02	0.04	0.06
\hat{p}_1	0.412	0.428	0.441	0.43	0.422	0.416	0.403

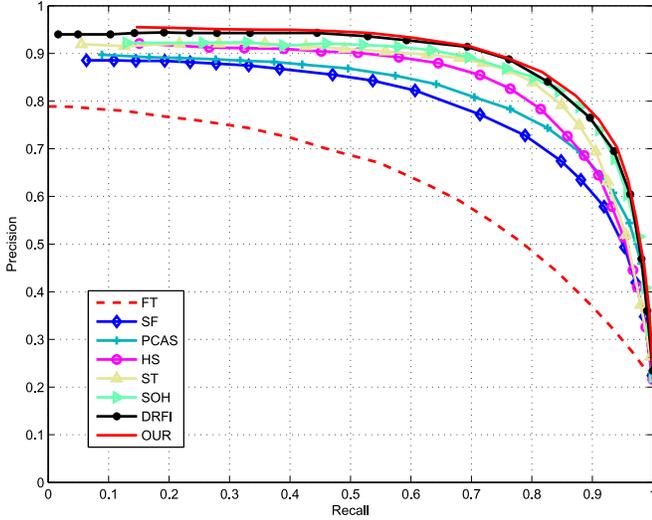


Fig. 10. P-R curves of the different methods on the MSRA10K dataset.

1) *Performance on the MSRA10K Dataset:* Fig. 10 illustrates the comparison results of the different methods by P-R curves, where the proposed method is denoted as “OUR” for short. It can be seen that the proposed method, as well as DRFI, ST, and SOH, significantly outperform the remaining methods. Furthermore, the HS [17] achieves better performance than the PCAS [28] and filter saliency [47]. In general, the Precision of these methods decreases over the Recall. In the P-R curves, it can be seen that the precision is quickly decreasing when the recall is near 0.9 because the false salient object masks obtained by thresholding the saliency map increase sharply.

Fig. 11 shows the performance of these methods in terms of Precision, Recall, F -measure, and MAE, where F -measure shows their overall performance. It can be seen that the proposed method has achieved the best performance in terms of MAE and F -measure.

In Fig. 12, we demonstrated the examples for the image saliency detection, in which the brighter areas correspond to the more salient regions of the image. It can be seen that the proposed method can highlight the whole salient regions and recognize the appearance of the salient object. In contrast, the PCAS method [28] usually highlights the image edges, but not the whole salient regions. The result of SF [47] is not as distinct as that in our salient region. The ST [35] and SOH [36] can detect more saliency, as well as DRFI [18].

As for the DRFI [18], it demonstrates the satisfactory performance in terms of MAE and F -measure, but the appearances of the salient objects are usually with some noisy pixels. In fact, both of the DRFI and the proposed method try to extract the local visual attributes, which jointly determine the visual saliency. Besides that, the proposed method learns the

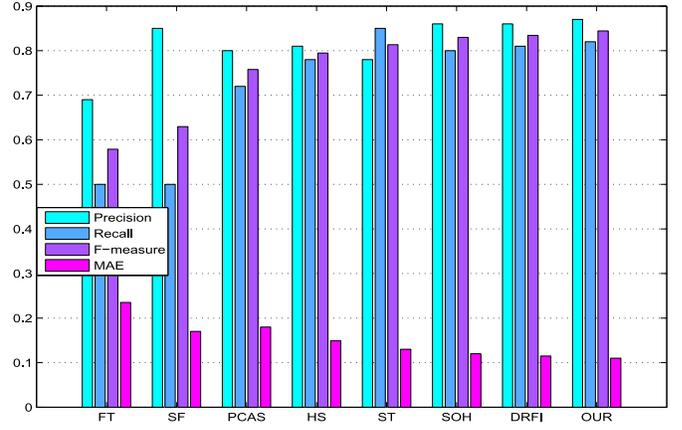


Fig. 11. Performance of the different methods on the MSRA10K dataset.

weight for the four complementary saliency maps and adjusts the weights to integrate these maps to achieve an improved saliency.

2) *Performance on the ECSSD Dataset:* In this experiment, we performed the saliency detection on the ECSSD dataset. The P-R curves are plotted in Fig. 13, which shows the results of saliency detection by these different methods. From this figure, we can see that SOH [36], ST [35], DRFI [18], HS [17], and the proposed method give better performance in comparison with SF [47], PCAS [28], and FT [22] because the former algorithms adopt more prior knowledge or visual features in their models. Further, the performance of the proposed method is comparable to the DRFI [18], both of which are slightly better than the latest SOH, ST, and HS.

Also, Fig. 14 shows the comparison results in terms of Precision, Recall, MAE, and F -measure for each method. Actually, it is more difficult to maintain a high F -measure on the ECSSD dataset because the images in this dataset are usually with more complicated background that makes the segmentation more challenging. Once again, it can be seen that the proposed method outperforms the other ones in terms of MAE and F -measure.

3) *Discussion:* This section discusses the effect of the preprocessing (i.e., smoothing operation) on the saliency detection, and presents a further comparison of the contribution of each saliency component, i.e., the previously stated S_1 , S_2 , S_3 , and S_4 , in the proposed method.

The preprocessing procedure is capable of generating more homogeneous background that facilitates the detection of image saliency. Fig. 15 shows an example of the effect of smoothing operation on the local color saliency. It can be observed that the local color saliency is clearer than the one without the preprocessing. Thus, it contributes to the overall saliency detection. We have found that the precision is improved by about 2% using the preprocessing.

As for the contribution of each saliency component, Fig. 16 shows the comparison results of the proposed method associated with the different combinations of these four saliency components. It can be seen that the global color saliency is the most critical one in the proposed method, while the local color saliency is ranked the second. Also, the color attribute

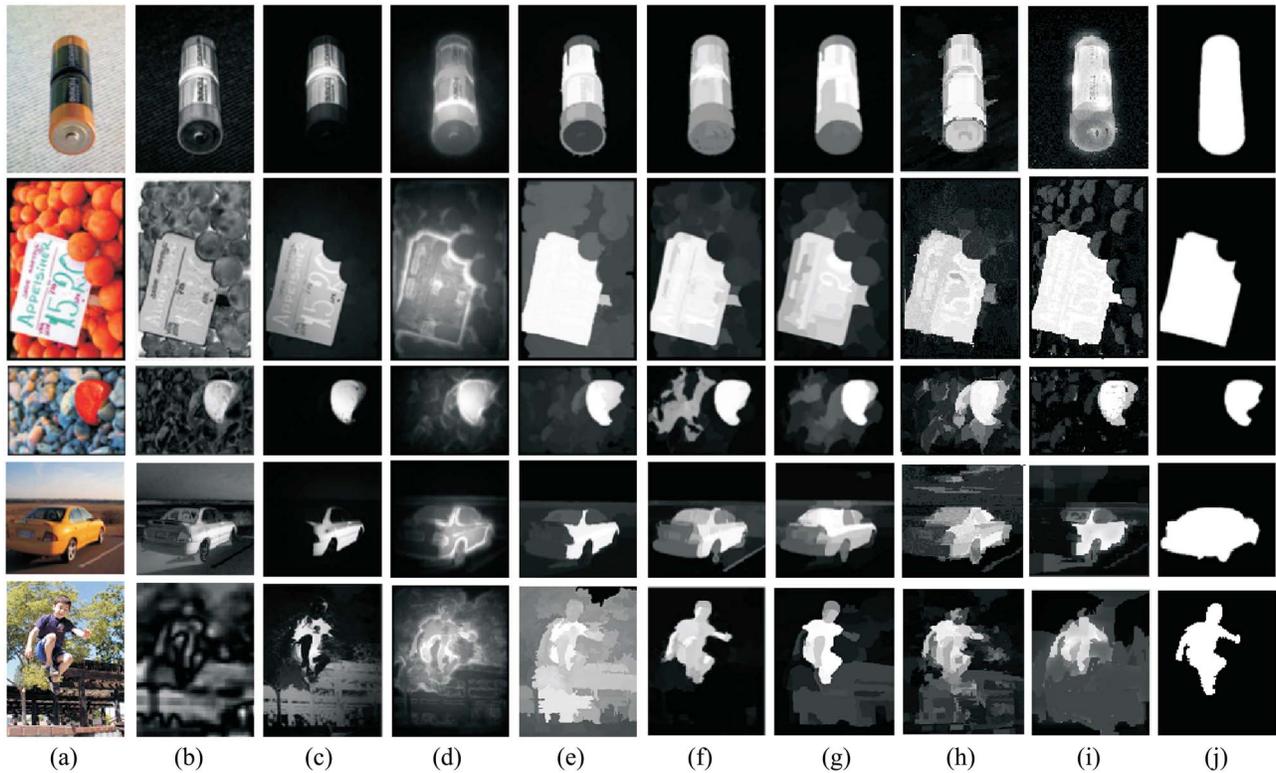


Fig. 12. Experimental results on some test images. (a) Input images. (b) FT [22]. (c) SF [47]. (d) PCAS [28]. (e) HS [17]. (f) ST [35]. (g) SOH [36]. (h) DRFI [18]. (i) Proposed method. (j) Ground truth.

TABLE II
AVERAGE RUNNING TIME OF THE PROPOSED METHOD AND ITS COUNTERPARTS (SECONDS PER IMAGES)

Method	FT [22]	SF [47]	PCAS [28]	HS [17]	ST [35]	SOH [36]	DRFI [18]	OUR
Time (s)	0.16	5.2	2.3	2.5	67.8	12.7	36.4	9.2
Code & C++	Matlab	Matlab	Matlab & C++	Matlab & C++	Matlab & C++	Matlab	Matlab & C++	Matlab

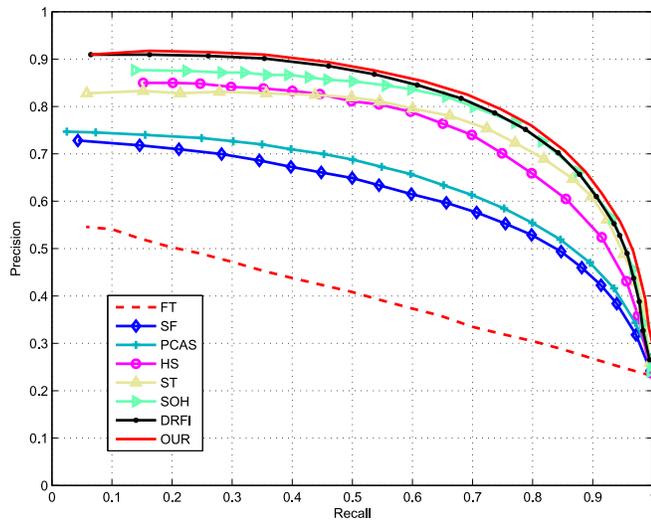


Fig. 13. P-R curves of the different methods on the ECSSD dataset.

plays a key role in the human vision system, especially the global color cue. As for the global edge saliency, the edge feature is ranked the third as salient objects tend to have obvious boundary in most images. In contrast, the local textural

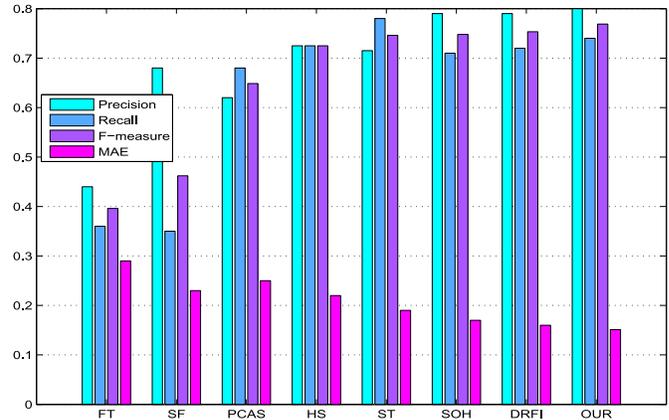


Fig. 14. Performance of the different methods on the ECSSD dataset.

saliency, which describes the local details in some sense, is not as crucial as the other components in all cases we have tried thus far. Nevertheless, when the textural patterns in the image are obvious, e.g., some remote sensing images [48], the textural saliency will play an important role as well.

Finally, we evaluated the effectiveness of the proposed approach on the low-resolution image with multiple objects

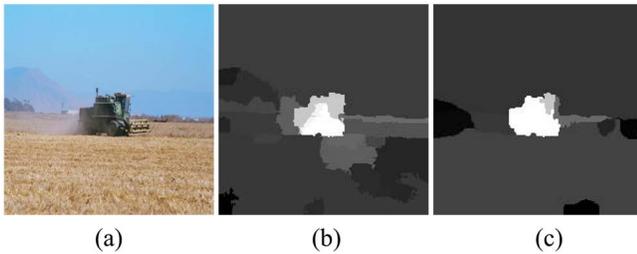


Fig. 15. Effect of preprocessing on the local color saliency, where (a) input image, (b) local color saliency without preprocessing, and (c) local color saliency using preprocessing.

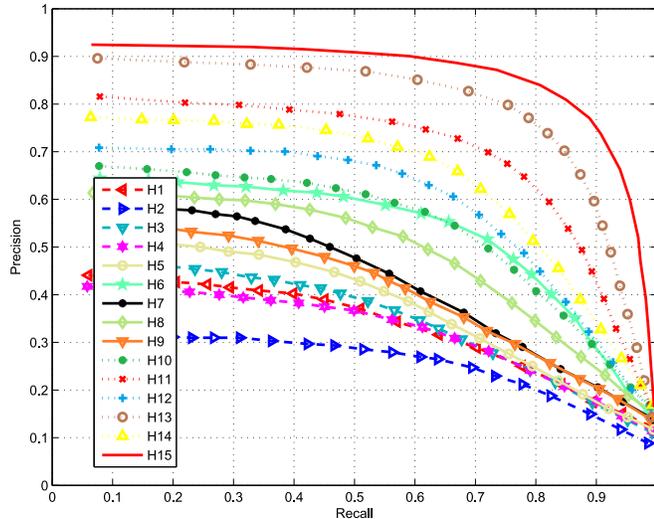


Fig. 16. Comparison results of the proposed method with the different saliency combinations: $H1 = \{S_1\}$, $H2 = \{S_2\}$, $H3 = \{S_3\}$, $H4 = \{S_4\}$; and $H5 = \{S_1, S_2\}$, $H6 = \{S_1, S_3\}$, $H7 = \{S_1, S_4\}$, $H8 = \{S_2, S_3\}$, $H9 = \{S_2, S_4\}$, $H10 = \{S_3, S_4\}$; and $H11 = \{S_1, S_2, S_3\}$, $H12 = \{S_1, S_2, S_4\}$, $H13 = \{S_1, S_3, S_4\}$, $H14 = \{S_2, S_3, S_4\}$, $H15 = \{S_1, S_2, S_3, S_4\}$.

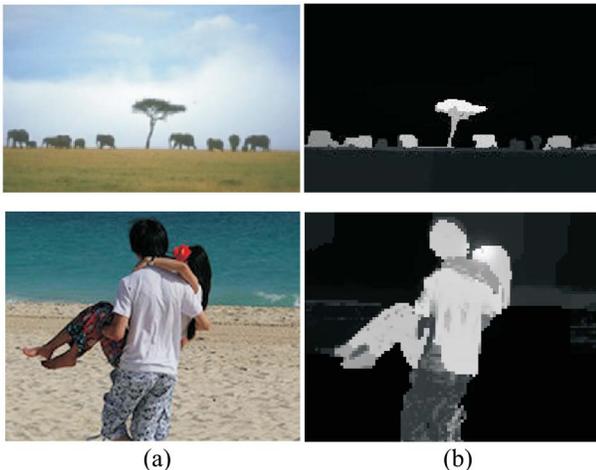


Fig. 17. Example of multiple saliency detection, in which (a) is the input image and (b) is the saliency detection result.

and image with clustered objects, respectively, as shown in Fig. 17. Once again, it can be seen that the proposed approach is able to detect the image saliencies successfully in these cases we have tried so far. This implies that the proposed

local and global features, as well as its combination scheme, are effective for the saliency detection task.

4) *Running Time*: The average computation time of these methods running at a machine with an Intel Dual Core i7-3770 3.40 GHz CPU is shown in Table II. It can be seen that the running time of the proposed method is faster than ST, SOH, and DRFI, but not as fast as FT, SF, PCAS, and HS. In fact, most time consumed by the proposed method is at the preprocessing step. One feasible way to further speed up the proposed method is to implement it using C language with the optimal codes, as well as the GPU technique. We will leave it elsewhere in our future work.

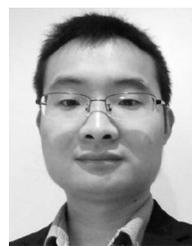
V. CONCLUSION

In this paper, we have proposed a novel method to detect image saliency by utilizing both local and global features. First, we have proposed a variant of smooth scheme based on the global minimization algorithm to smooth the image, featuring the automatic selection of smoothing parameter. After partitioning the smoothed image into a set of regions, we compute the local color saliency in the segmented regions, and then estimate the local textural saliency. Furthermore, we have adopted the global color distribution embedded with color coherence, and utilized the multiple salient edge to calculate the global saliency. Subsequently, we combine the local and global saliency, and utilize the composition information to perform postprocessing so that the final saliency is obtained. Compared to the existing counterparts, the proposed method features: 1) the enhanced accuracy of detecting visual salient region and appearance, 2) the robustness against the noise and the low-resolution problem of images to a great extent, and 3) its applicability to detecting multiple saliency detection task. In the future, we will extend this model to the video saliency detection as in [49].

REFERENCES

- [1] L. Itti and C. Koch, "Computational modelling of visual attention," *Nat. Rev. Neurosci.*, vol. 2, no. 3, pp. 194–203, 2001.
- [2] B. P. Ölveczky, S. A. Baccus, and M. Meister, "Segregation of object and background motion in the retina," *Nature*, vol. 423, no. 6938, pp. 401–408, 2003.
- [3] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition?" in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Washington, DC, USA, 2004, pp. 37–44.
- [4] A. P. Bradley and F. W. M. Stentiford, "Visual attention for region of interest coding in JPEG 2000," *J. Vis. Commun. Image Represent.*, vol. 14, no. 3, pp. 232–250, 2003.
- [5] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1304–1318, Oct. 2004.
- [6] A. Santella, M. Agrawala, D. DeCarlo, D. Salesin, and M. Cohen, "Gaze-based interaction for semi-automatic photo cropping," in *Proc. SIGCHI Conf. Human Fact. Comput. Syst.*, Montreal, QC, Canada, 2006, pp. 771–780.
- [7] Y. Fang, Z. Chen, W. Lin, and C.-W. Lin, "Saliency detection in the compressed domain for adaptive image retargeting," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 3888–3901, Sep. 2012.
- [8] Z. Yücel *et al.*, "Joint attention by gaze interpolation and saliency," *IEEE Trans. Cybern.*, vol. 43, no. 3, pp. 829–842, Jun. 2013.
- [9] A. Oikonomopoulos, I. Patras, and M. Pantic, "Spatiotemporal salient points for visual recognition of human actions," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 36, no. 3, pp. 710–719, Jun. 2006.

- [10] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu, "Sketch2Photo: Internet image montage," *ACM Trans. Graph.*, vol. 28, no. 5, pp. 1–10, 2009.
- [11] T. N. Vikrma, M. Tscherepanow, and B. Wrede, "A saliency map based on sampling an image into random rectangular regions of interest," *Pattern Recognit.*, vol. 45, no. 9, pp. 3114–3124, 2012.
- [12] G. Kootstra, A. Nederveen, and B. De Boer, "Paying attention to symmetry," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Leeds, U.K., 2008, pp. 1115–1125.
- [13] R. Valenti, N. Sebe, and T. Gevers, "Image saliency by isocentric curvedness and color," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Kyoto, Japan, 2009, pp. 2185–2192.
- [14] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Kyoto, Japan, 2009, pp. 2106–2113.
- [15] J. Li, D. Xu, and W. Gao, "Removing label ambiguity in learning-based visual saliency estimation," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1513–1525, Apr. 2012.
- [16] N. Bruce and J. Tsotsos, "Saliency based on information maximization," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 18. Vancouver, BC, Canada, 2006, pp. 155–162.
- [17] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Portland, OR, USA, 2013, pp. 1155–1162.
- [18] H. Jiang *et al.*, "Salient object detection: A discriminative regional feature integration approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Portland, OR, USA, 2013, pp. 2083–2090.
- [19] P. L. Rosin, "A simple method for detecting salient regions," *Pattern Recognit.*, vol. 42, no. 11, pp. 2363–2371, 2009.
- [20] M. Holtzman-Gazit, L. Zelnik-Manor, and I. Yavneh, "Salient edges: A multi scale approach," in *Proc. ECCV Workshop Vis. Cogn. Tasks*, vol. 10. Heraklion, Greece, Sep. 2010, p. 4310.
- [21] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Minneapolis, MN, USA, 2007, pp. 1–8.
- [22] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Miami, FL, USA, 2009, pp. 1597–1604.
- [23] N. Murray, M. Vanrell, X. Otazu, and C. Parraga, "Saliency estimation using a non-parametric low-level vision model," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Providence, RI, USA, 2011, pp. 433–440.
- [24] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 185–198, Jan. 2010.
- [25] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 19. Vancouver, BC, Canada, 2007, pp. 545–552.
- [26] T. Avraham and M. Lindenbaum, "Esaliency (extended saliency): Meaningful attention using stochastic image modeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 4, pp. 693–708, Apr. 2010.
- [27] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Providence, RI, USA, 2011, pp. 409–416.
- [28] R. Margolin, A. Tal, and L. Zelnik-Manor, "What makes a patch distinct?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Portland, OR, USA, 2013, pp. 1139–1146.
- [29] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
- [30] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Francisco, CA, USA, 2010, pp. 2376–2383.
- [31] D. Gao and N. Vasconcelos, "Decision-theoretic saliency: Computational principles, biological plausibility, and implications for neurophysiology and psychophysics," *Neural Comput.*, vol. 21, no. 1, pp. 239–271, 2009.
- [32] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cogn. Psychol.*, vol. 12, no. 1, pp. 97–136, 1980.
- [33] J. Li, M. D. Levine, X. An, and H. He, "Saliency detection based on frequency and spatial domain analysis," in *Proc. Brit. Mach. Vis. Conf.*, Dundee, U.K., 2011, pp. 1–11.
- [34] T. Liu *et al.*, "Learning to detect a salient object," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 353–367, Feb. 2011.
- [35] Z. Liu, W. Zou, and O. Le Meur, "Saliency tree: A novel saliency detection framework," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 1937–1952, May 2014.
- [36] V. Vilaplana, "Saliency maps on image hierarchies," *Signal Process. Image Commun.*, vol. 38, pp. 84–99, Oct. 2015.
- [37] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via L_0 gradient minimization," *ACM Trans. Graph.*, vol. 30, no. 6, pp. 1–11, 2011.
- [38] M.-M. Cheng *et al.*, "Efficient salient region detection with soft image abstraction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Sydney, NSW, Australia, 2013, pp. 1529–1536.
- [39] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
- [40] R. Lopes *et al.*, "Local fractal and multifractal features for volumic texture characterization," *Pattern Recognit.*, vol. 44, no. 8, pp. 1690–1697, 2011.
- [41] A. Çarkacıoğlu and F. T. Yarman-Vural, "SASI: A generic texture descriptor for image retrieval," *Pattern Recognit.*, vol. 36, no. 11, pp. 2615–2633, 2003.
- [42] T. Ojala, M. Pietikainen, and T. Maenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
- [43] H. Tian *et al.*, "Salient region detection by fusing bottom-up and top-down features extracted from a single image," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4389–4398, Oct. 2014.
- [44] O. F. Lazareva, E. A. Wasserman, and I. Biederman, "Pigeons and humans are more sensitive to nonaccidental than to metric changes in visual objects," *Behav. Process.*, vol. 77, no. 2, pp. 199–209, 2008.
- [45] T. Ahonen, J. Matas, C. He, and M. Pietikäinen, "Rotation invariant image description with local binary pattern histogram Fourier features," in *Proc. 16th Scandinavian Conf. Image Anal.*, Oslo, Norway, 2009, pp. 61–70.
- [46] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, Mar. 2015.
- [47] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Providence, RI, USA, 2012, pp. 733–740.
- [48] L. Zhang, J. Chen, and B. Qiu, "Region of interest extraction in remote sensing images by saliency analysis with the normal directional lifting wavelet transform," *Neurocomputing*, vol. 179, pp. 186–201, Feb. 2016.
- [49] Y. Fang, W. Lin, Z. Chen, C.-M. Tsai, and C.-W. Lin, "A video saliency detection model in compressed domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 1, pp. 27–38, Jan. 2014.



Qinmu Peng received the Ph.D. degree from Hong Kong Baptist University, Hong Kong, in 2015.

His current research interests include image processing, pattern recognition, and machine learning methods in computer vision.

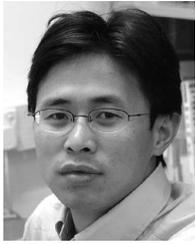


Yiu-ming Cheung (SM'06) received the Ph.D. degree from the Chinese University of Hong Kong, Hong Kong.

He is a Full Professor with the Department of Computer Science, Hong Kong Baptist University, Hong Kong. His current research interests including machine learning, pattern recognition, visual computing, and optimization.

Prof. Cheung is the Founding Chairman of the Computational Intelligence Chapter of the IEEE Hong Kong Section. He is currently serving as an

Associate Editor for the *IEEE Transactions on Neural Networks and Learning Systems*, *Knowledge and Information Systems*, and the *International Journal of Pattern Recognition and Artificial Intelligence*. He is a Senior Member of ACM.



Xinge You (M'08–SM'10) received the B.S. and M.S. degrees in mathematics from Hubei University, Wuhan, China, in 1990 and 2000, respectively, and the Ph.D. degree from Hong Kong Baptist University, Hong Kong, in 2004.

He is currently a Professor with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan. His research interests include wavelets and its application, signal and image processing, pattern recognition, machine learning, and computer vision.



Yuan Yan Tang (F'04) received the B.S. degree in electrical and computer engineering from Chongqing University, Chongqing, China, the M.Eng. degree in electrical engineering from the Beijing Institute of Post and Telecommunications, Beijing, China, and the Ph.D. degree in computer science from Concordia University, Montreal, QC, Canada.

He is currently the Chair Professor with the Faculty of Science and Technology, University of Macau, Macau, China, and a Professor/Adjunct Professor/Honorary Professor with several institutes,

including several universities in China, Concordia University, and Hong Kong Baptist University, Hong Kong. He has published over 400 technical papers and authored/coauthored over 25 monographs/books/bookchapters on subjects ranging from electrical engineering to computer science. His current research interests include wavelet theory and applications, pattern recognition, image processing, document processing, artificial intelligence, and Chinese computing.

Dr. Tang is the Founder and an Editor-in-Chief of the *International Journal on Wavelets, Multiresolution, and Information Processing* and an Associate Editor of several international journals, such as the *International Journal on Pattern Recognition and Artificial Intelligence*. He is the Founder and the Chair of the Pattern Recognition Committee in the IEEE Systems, Man, and Cybernetics Society. He has served as the General Chair, the Program Chair, and a Committee Member for many international conferences, including the General Chair of the 18th International Conference on Pattern Recognition. He is the Founder and the General Chair of the series International Conferences on Wavelets Analysis and Pattern Recognition. He is a Fellow of the International Associate of Pattern Recognition.