Exploiting color volume and color difference for salient region detection

Guang-Hai Liu and Jing-Yu Yang *Member, IEEE*

Abstract—Foreground and background cues can assist humans in quickly understanding visual scenes. In computer vision, however, it is difficult to detect salient objects when they touch the image boundary. Hence, detecting salient objects robustly under such circumstances without sacrificing precision and recall can be challenging. In this study, we propose a novel model for salient region detection, namely, the foreground-center-background (FCB) saliency model. Its main highlights are: 1) we use regional color volume as the foreground, together with perceptually uniform color differences within regions to detect salient regions. This can highlight salient objects robustly, even when they touched the image boundary, without greatly sacrificing precision and recall. 2) We employ center saliency to detect salient regions together with foreground and background cues, which improves saliency detection performance. 3) We propose a novel and simple yet efficient method that combines foreground, center, and background saliency. Experimental validation with three well-known benchmark datasets indicates that the FCB model outperforms several state-of-the-art methods in terms of precision, recall, F-measure and, particularly, the mean absolute error. Salient regions are brighter than those of some existing state-of-the-art methods.

Index Terms— Visual attention, L*a*b* color space, regional color volume, perceptually uniform color difference, saliency detection

1 Introduction

Saliency detection can be broadly defined as the accurate visual detection of objects that warrant attention. Humans pay greater attention to the regions of their visual field that have higher contrast than neighboring regions. Many applications, such as fixation prediction [1], object recognition [2], image segmentation [3], and content-based image retrieval [4] can benefit from saliency detection. Existing state-of-the-art saliency detection methods are broadly categorized into bottom-up vs top-down models, and global contrast vs local contrast-based methods.

In recent years, foregrounds and backgrounds have been widely and effectively utilized for salient object detection [13], [14], [15], [16], [17], [18], [19], [27]. Four image boundary regions are generally considered in boundary contrast calculations, whereas foreground regions are similar in terms of having coherent and consistent visual appearances [13], [14], [15], [16], [17], [18], [19]. However, it is difficult to detect salient objects when they touch the image boundary. Some examples of this problem are shown in Fig. 1. Methods of robustly detecting salient objects under such circumstances without sacrificing precision and recall can be challenging. To address this problem, we propose a novel and very simple yet efficient saliency model named the foreground-center-background (FCB) saliency model. It works by exploiting color volume and color differences within regions of the L*a*b* color space.

The major steps in the FCB model are as follows: first, the input image is segmented into regions by a simple linear iterative clustering (SLIC) algorithm [42], and the regional color volume and perceptually uniform color differences within the regions are computed, with the regional color volume considered as the foreground. Second, we calculate the background saliency and center saliency maps. Finally, we integrate the foreground saliency, center saliency, and background saliency.

The FCB model can be summarized in three parts. The first part uses a method we call *regional color volume* to define the foreground. It highlights salient objects robustly, even when they touch the image boundaries, without sacrificing precision and recall greatly. The second part, *center saliency*, detects salient regions together with foreground and background cues, which improves the performance of saliency detection. Third, a novel and simple yet efficient method is proposed that combines the foreground, center, and background saliencies, whereby the salient regions detected are often brighter than those of some existing state-of-the-art methods.

The remainder of the present study is organized as follows. In Section 2, existing state-of-the-art saliency detection methods are reviewed. The proposed saliency model is presented in Section 3. In Section 4, performance comparisons are made with three benchmark datasets. Section 5 concludes the study.

- *G.-H. Liu is with the College of Computer Science and Information Technology, Guangxi Normal University, Guilin 541004, China. E-mail: liuguanghai009@163.com*
- *J.-H. Yang is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China. E-mail yangjy@mail.njust.edu.cn*

Fig. 1. Some examples of salient objects touching image boundaries. The top row of images shows the original images, while the bottom row shows the corresponding ground-truthed images.

2 RELATED WORKS

Many methods for detecting salient objects and regions have been presented. There are a variety of visual saliency techniques, with fixation prediction and salient region detection being two of the major branches.

2.1 Eye fixation prediction

In the early days, fixation prediction was an important branch of saliency detection. It exploits the contrast of low-level features, such as color, orientation, and intensity.

In 1998, Itti et al. [1] introduced central-surrounded contrast to detect saliency using color, intensity, and orientation features. Ma and Zhang [5] presented a new saliency map generation method based on local contrast analysis. Based on current understanding of the human visual system's behavior, Le Meur et al. [6] presented a coherent computational approach to the modeling of bottom-up visual attention. Borji and Itti [7] introduced a saliency model that considers local and global image patch rarities as two complementary processes. Harel et al. [8] proposed a graph-based (GB) visual saliency scheme to detect salient regions. Frintrop et al. [9] adapted Itti et al.'s [1] saliency model by introducing a twin pyramid for computing Gaussian differences and utilized it for salient object segmentation.

2.2 Salient object detection

Salient object detection methods based on global features have demonstrated advantages in some tasks. Salient object detection considers saliency detection as a binary segmentation problem. It can be broadly categorized into bottom-up and top-down models.

2.2.1 Bottom-up models

Bottom-up models only use low-level features for saliency detection. Achanta et al. [3] introduced a segmentation method for salient region detection. Liu et al. [4] calculated the local contrast of edges and color volume in the HSV color space to detect salient regions and used further processing compared to traditional content-based image retrieval (CBIR). Achanta et al. [10] introduced a frequency-tuned (FT) method for salient region detection. It exploits color and luminance features and can output full resolution saliency maps with well-defined boundaries for salient objects. Goferman et al. [2] presented a context-aware (CA) saliency detection method which aims to detect the image regions that represent the scene. Cheng et al. [11] [12] [29] presented a histogram-based contrast (HC) method and utilized global cues (GC) and region-based contrast (RC) for saliency detection. Meanwhile, Perazzi et al. [30] proposed the use of saliency filters (SF) for salient region detection. Kim et al. [24] introduced a novel technique that automatically detects the salient regions of an image via high-dimensional color transform (HDCT). Zhang et al. [31] proposed a novel Boolean map-based saliency (BMS) model. Li et al. [27] explored the problem of utilizing light fields for saliency detection, as they provide useful focus, depth, and abjectness cues.

2.2.2 Top-down models

Top-down models utilize high-level priors or task-specific learning algorithms for saliency detection.

High-level priors. Wei et al. [13] proposed a geodesic saliency detection method that benefits from background priors. Jiang et al. [14] jointly considered the divergence and spatial distribution of salient objects and the background. Zhu et al. [15] presented a novel background measurement that uses intuitive and clear geometrical interpretation for high accuracy background detection and saliency estimation (RBD). Jiang et al. [16] proposed a novel segmentation algorithm to integrate bottom-up salient stimuli and object-level shape priors. Zhang et al. [17] proposed an approximate minimum barrier distance (MBD) transform algorithm for salient object detection. Yang et al. [18], [19] presented a graph-based manifold ranking (GMR) method to extract background regions and foreground salient objects and extended it to a multi-scale graph. Yan et al. [20] proposed a hierarchical model (HS) to reduce the effect of complex structures, while Shi [21] extended the hierarchical model to utilizing a multi-layer approach. Shen et al. [25] adopted low rank (LR) matrix recovery that incorporates traditional low-level features with higher-level guidance to detect salient objects. Peng et al. [26] proposed a novel structured matrix decomposition (SMD) model with two structural regularizations for salient object detection. Liu et al. [32] proposed a saliency tree (ST) model for effective saliency measurements.

Task-specific learning algorithms: Margolin et al. [22] proposed a novel approach to detecting salient regions via identifying unique patterns (PCA), while Hou et al. [23] proposed a spectral residual (SR) approach. Wang et al. [28] [49] presented supervised feature integration to discriminatively integrate salient features for saliency computation. He et al. [33] proposed a novel method that uses a superpixel-wise convolutional neural network (CNN) to detect salient regions. Zhao et al. [34] utilized multi-content deep learning techniques to detect saliency, whereby global and local contexts are both taken into account. Chen et al. [35] utilized the deep CNN method to learn saliency representations in a progressive manner, and Li et al. [36] proposed a multi-task deep saliency model based on a fully-CNN. Li [37] proposed an end-to-end deep contrast network for salient object detection utilizing deep learning techniques. Liu et al. [38] proposed a deep hierarchical saliency network for salient object detection.

The above-mentioned methods are some of the existing state-of-the-art saliency detection methods. Toet [39] and Borji and Itti [40] have presented reviews of existing saliency models.

However, creating a saliency detection method that has highly efficient computing performance and generates high-quality saliency maps remains challenging.

3 THE PROPOSED SALIENCY MODEL

The L*a*b* color space has good perceptual uniformity. The perceptually uniform color difference between two color points can be measured as a Euclidean distance that is close to human color perception [44] [45]. Color volumes derived from the ellipsoid shape of the L*a*b* color space contain rich visual information. The attributes of the L*, a* and b* components can be integrated into a new attribute such that the conspicuously colored areas are highlighted while some backgrounds are suppressed. Thus, color volume, together with perceptually uniform color differences, are the basis of the present study.

In [4], we proposed the concept of *color volume* derived from the shape of the HSV color space. Color volume and edges are used to detect salient regions in a similar manner to Itti et al.'s attention model [1]. Then, visual features are encoded as representations and used for CBIR. In our prior work [41], the same concept of color volume was also utilized to detect salient regions in a similar manner to Itti et al.'s attention model [1]. However, only utilizing saliency based on color volume has a major drawback; if the difference between the foreground and background colors is small, then saliency detection is poor.

In order to overcome this shortcoming, the present study combines the approach of defining the regional color volume with a new approach of detecting perceptually uniform color differences within regions. This can enhance saliency detection.

3.1 Foreground Saliency

Generally, an object in the closer area of the visual field can be considered to be the foreground. In the present study, regional color volume can easily highlight the salient objects in many cases, thus we utilized it as the foreground saliency method in the L*a*b* color space.

The L*a*b* color space can be simply interpreted as an ellipsoid, as shown in Fig. 2. It is excellent for extracting features that utilize the volume of the ellipsoid shape, termed the *color volume*.

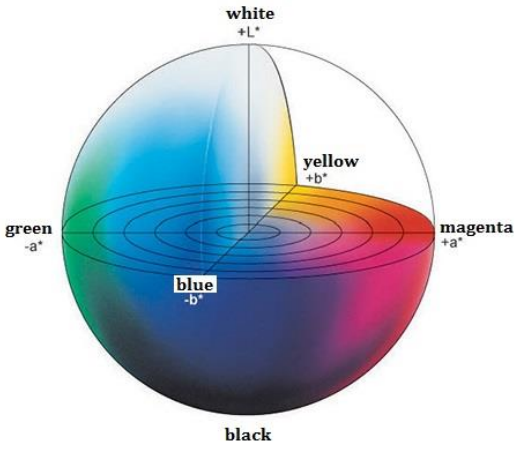


Fig. 2. The ellipsoid shape for L*a*b* color space (Source: [46]).

In the L*a*b* color space, the L* component represents the lightness of the color, i.e. L* = 0 is black and L* = 100 indicates diffuse white, while specular white may have a higher value. The position of the a* component is between red/magenta and green, with negative a* values indicating green and positive a* values indicating red/magenta. The position of b* is between yellow and blue, with negative b* values indicating blue and positive b* values indicating yellow [47]. For a random dot (L, a, b) in the L*a*b* color space, its color volume can be defined as:

$$cv_{Lab}(x,y) = \frac{4}{3}\pi \times L(x,y) \times a(x,y) \times b(x,y) \quad (1)$$

In the present study, we adopted SLIC for regional segmentation in the L*a*b* color space. Let there be an $W \times H$ input image f. We segment it into n regions applying the SLIC algorithm [42], and the segmented regions are denoted as $R_1, R_2, \ldots, R_{n,}$. The average perceptually uniform color components of regions R_i are calculated as follows:

$$\mathcal{L}_i = \sum_{x,y \in R_i} L(x,y)/count(x,y) \quad (2)$$

$$\mathcal{A}_i = \sum_{x,y \in R_i} a(x,y)/count(x,y) \quad (3)$$

$$\mathcal{B}_i = \sum_{x,y \in R_i} b(x,y)/count(x,y) \quad (4)$$

In the above equations, x and y are the pixel coordinates of region R_i and $count(x,y)$ denotes the pixel number in region R_i. Based on the color components $\mathcal{L}_{\mathrm{i}}$, $\mathcal{A}_{\mathrm{i}}$, and $\mathcal{B}_{\mathrm{i}}$, the regional color volume ($\mathcal{RCV}$) can be defined as:

$$\mathcal{RCV}(i) = \frac{4}{3}\pi \times \mathcal{L}_i \times \mathcal{A}_i \times \mathcal{B}_i \quad (5)$$

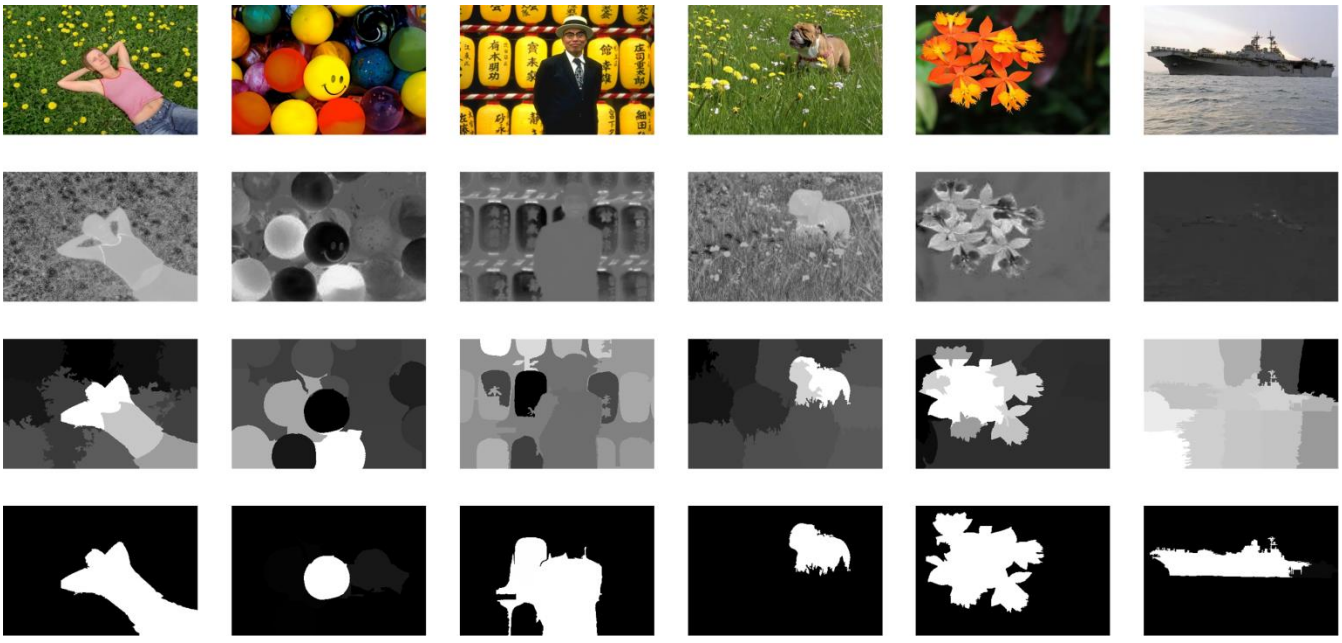


Fig. 3. Color volume maps of the L*a*b* color space. Row 1: input images, row 2: maps of color volume cv_{Lab}, row 3: maps of regional color volume using the SLIC algorithm for regional segmentation (number of regions = 20), row 4: final saliency maps.

Examples of cv_{Lab} and $\mathcal{RCV}$ are shown in Fig. 3. As can be seen, by utilizing cv_{Lab} and $\mathcal{RCV}$, the conspicuous color areas and major objects are highlighted, while some backgrounds are suppressed.

Far more important is that cv_{Lab} and $\mathcal{RCV}$ can handle salient objects that touch the image boundary, which is one of the highlights of our method. For instance, in columns 1 and 3 of Fig. 3, there are images of a woman lying on grass and a man standing in front of a group of lanterns, both of whom are touching the image boundaries. Utilizing cv_{Lab} and $\mathcal{RCV}$ can preserve salient objects that touch the image boundary. Compared to cv_{Lab}, the conspicuity regions or the salient objects are enhanced by utilizing $\mathcal{RCV}$.

3.2 Background Saliency

In some state-of-the-art methods [17], [18], [19], [20], four image boundary regions (upper, lower, left, and right) are considered during boundary contrast calculations, whereas foreground regions are considered similar in terms of having a coherent and consistent visual appearance. In these methods, the background regions are usually considered to be similar to one of the four image boundaries in terms of local or global appearance.

However, if the salient objects touch the image boundary, then these methods may fail. In some cases, the image boundary regions also belong to the salient objects. To reduce or address this problem, we consider all image boundary regions in the calculation of the background. An example is shown in Fig. 4.

Let there be an W × H input image f, we segment it into $m + n$ regions applying the SLIC algorithm, and denote the segmented regions as $R_1, R_2, \dots, R_m, R_{m+1}, \dots, R_{m+n}$. The first m regions are the boundary regions and the remaining n regions are the no-boundary regions.

(a) (b) (c)

Fig. 4. Consideration of image boundary regions in our method. (a) Input image. (b) Image boundary regions (red circles) are segmented using the SLIC algorithm (number of regions = 20). (c) Saliency map.

In the SLIC algorithm, the pixel coordinates are normalized to [0, 1]. Here, we define the background saliency for region R_i in an image with a total of $m+n$ regions. Details of the calculations of background saliency are as follows.

First, the sum value of saliency between regions R_i and all other boundary regions $R_j, j = 1,2,..,m$ is calculated and denoted as $B(i), i = 1,2,\dots,m+n$:

$$B(i) = \sum_{j=1}^{m}\left\{ w_{ij}^{d} w_i^{c} \left\| C_i - C_j \right\|^2 \right\}, \quad i = 1,2,\dots,m+n;\ j = 1,2,..,m \tag{6}$$

$$w_{ij}^{d} = 1.0 - \sqrt{\frac{(x_i - x_j)^2 + (y_i - y_j)^2}{2.0}} \tag{7}$$

$$w_i^{c} = 1.0 - \sqrt{\frac{(x_i - 0.5)^2 + (y_i - 0.5)^2}{0.5}} \tag{8}$$

In (6) and (9), $\left\| C_i - C_j \right\|$ is the Euclidean distance between average colors $C_i = \{\mathcal{L}_i, \mathcal{A}_i, \mathcal{B}_i\}$ and $C_j = \{\mathcal{L}_j, \mathcal{A}_j, \mathcal{B}_j\}$, and w_{ij}^{d} and w_i^{c} denote the spatial weight and center bias weight, respectively. w_{ij}^{d} is utilized to increase the effects of closer regions and decrease the effects of more distant regions. If region R_i is close to the center of the input image, w_i^{c} has a high value, and vice versa.

Second, the maximum saliency between regions R_i and all other boundary regions $R_j, j = 1,2,..,m$ is calculated and denoted as $M(i), i = 1,2,\dots,m+n$:

$$M(i) = max_{\{j=1,..,m\}}\left\{ w_{ij}^{d}\, w_i^{c} \left\| C_i - C_j \right\|^2 \right\}, \quad i = 1,2,\dots,m+n \tag{9}$$

The parameters C_i and C_j are the average perceptual uniformity colors of regions R_i and R_j, respectively. In the calculation of saliency, the centroid coordinates x_i and y_i of regions R_i and R_j are normalized to [0, 1].

Third, we calculate the background saliency map ($\mathcal{BSM}$). In order to increase detection stability, we adjust the background saliency scores to avoid them becoming too large, as follows:

$$BG(i) = B(i) - 2M(i) \tag{10}$$

In the above equation, adjusting the background saliency scores may result in them becoming too small. Scores that are too large or too small will not improve detection stability; thus, we control the scores of the background saliency map as follows:

$$\mathcal{BSM}(i) = max\{BG(i), mean\{BG\}\} \tag{11}$$

where $max\{.\}$ denotes the maximum value, and $mean\{.\}$ denotes the mean value.

3.3 Center Saliency

In background saliency detection, both the spatial weight and center bias weight were considered; however, the saliency between the center region R_c and other regions $R_i, i = 1,2,\dots,m+n$ is unknown. In saliency detection, a salient object is more likely to lie in the central region surrounded by the background [17]; however, such an assumption usually ignores the relationship between the center region and surrounding regions, and results in salient objects that touch the image boundary being ignored.

In the present study, we propose a new concept, *center saliency*, to express the saliency score between regions $R_i, i = 1,2,\dots,m+n$ and the center region $\mathrm{R_c}$. The center saliency map ($\mathcal{CSM}$) is defined as:

$$\mathcal{CSM}(i) = (1 - w_i^{c})^6 \left\| C_i - C_c \right\| \tag{12}$$

where $\left\| C_i - C_c \right\|$ is the Euclidean distance between average colors $C_i = \{\mathcal{L}_i, \mathcal{A}_i, \mathcal{B}_i\}$ and $C_c = \{\mathcal{L}_c, \mathcal{A}_c, \mathcal{B}_c\}$. A weight coefficient $(1 - w_i^{c})^6$ is utilized to increase the effect of regions that are more distant from the center region, and decrease the effects of closer regions. If the regions R_i were close to the center region R_c, the saliency score has a small value. This is the opposite of the center bias weight w_i^{c}.

3.4 Combined Saliency

Foreground, center, and background saliency all play very important roles in saliency detection; however, effectively combining them is a challenge. The values of $\mathcal{RCV}$, $\mathcal{CSM}$ and $\mathcal{BSM}$ were normalized into the range of [0, 1], then a saliency map $\mathcal{SM}$ was calculated according to two situations:

(1) If the value of $\mathcal{CSM}$ is equal to or greater than the value of $\mathcal{BSM}$, then the formula is as follows.

$$\mathcal{SM} = \begin{cases} 0.0 & if\ \ minv = 0.0 \\ \mathcal{BCD} - \mathcal{RCV} \times \mathcal{BSM} & else \end{cases} \tag{13}$$

(2) If the $\mathcal{CSM}$ value is less than the $\mathcal{BSM}$ value, then the following formula is used.

$$\mathcal{SM} = 2 \times \mathcal{BCD} - \mathcal{RCV} \times \mathcal{CSM} \tag{14}$$

The difference between $\mathcal{CSM}$ and $\mathcal{BSM}$ is calculated as $\mathcal{BCD} = \left\| \mathcal{CSM} - \mathcal{BSM} \right\|$. The minimum of the foreground, center, and background saliencies is defined as $minv = \mathrm{Min}\{\mathcal{RCV}, \mathrm{Min}\{\mathcal{CSM}, \mathcal{BSM}\}\}$. The maximum of the foreground, center, and background saliencies is defined as $maxv = \mathrm{Max}\{\mathcal{RCV}, \mathrm{Max}\{\mathcal{CSM}, \mathcal{BSM}\}\}$, and Max{.} and Mi n{.} denote the maximum and minimum values, respectively. To calculate the model saliency map S, we apply a Tanh function as an activation function, as follows:

$$S = Tanh\{6.0 \times \mathcal{SM}^6 + 1.0\} \tag{15}$$

In the present study, our method can handle salient objects that touch the image boundary. Finally, we normalize the values of S into the range [0, 1] and then evaluate the performance of this method in terms of precision/recall, F-measure, and mean absolute error (MAE).

4 EXPERIMENTAL VALIDATION OF SALIENT REGION DETECTION

The benchmark datasets, comparative methods, and performance metrics used to validate our approach are introduced in this section. In the present study, three popular benchmark datasets were adopted: ECSSD, DUT-OMRON, and MSRA10K. The MSRA10K dataset contains many images with a single salient object in front of a

background of objects that are mostly simple and smooth [51]. The ECSSD dataset contains images with more complex backgrounds [52]. The DUT-OMRON dataset consists of 5,168 high-quality images manually selected from more than 140,000 images [53]. For the evaluation of run time, the MSRA1000 dataset was used.

In this section, our method was compared to some of the current state-of-the-art methods, e.g., Itti et al.'s [1] saliency model (IT), CA [2], AC [3], GB [8], FT [10], HC [11], MBD [17], GMR [18], RS [19], HS [20], CHS [21], SR [23], HDCT [24], LR [25], SMD [26], RC [29], and ST [32]. In the proposed saliency detection model, the SLIC algorithm [42] was adopted to generate regions.

In the experiments, we slightly changed the C++ code of the original SLIC algorithm (available on the internet) and rewrote it in Visual C# 2013. In the setting of core parameters, the value of compactness was set as 20.0 and the final number of superpixels was 20. We evaluated the performance of these methods in terms of precision and recall curves, F-measure curves, and MAE metrics.

In the following sub-sections, evaluation metrics are introduced and saliency detection evaluations are conducted with the ECSSD, DUT-OMRON, and MSRA10K datasets.

4.1 Evaluation Metrics

The precision-recall (PR) curve, F-measure curve, and MAE are widely utilized as saliency evaluation metrics. Applying a threshold from 0 to 255, binarized maps of the model saliency map (S) can be generated and denoted as M, together with the ground truth (GT), and precision and recall can be calculated as:

$$precision = \frac{|M \cap GT|}{|M|} \tag{16}$$

$$recall = \frac{|M \cap GT|}{|GT|} \tag{17}$$

In various saliency evaluations, the F-measure is widely adopted for balancing precision and recall, and is defined as:

$$F_\beta = \frac{(1+\beta^2).precision.recall}{\beta^2.precision+recall} \tag{18}$$

Achanta et al. [10] proposed frequency-tuned salient region detection and recommended setting $\beta^2 = 0.3$ to emphasize precision rather than recall. Borji et al. [48] stated that the evaluation of non-salient regions can be unfair when utilizing overlapping-based evaluations, because the method successfully detects non-salient regions but misses salient ones. Therefore, MAE was used to evaluate performance, which can be calculated as follows:

$$MAE = \frac{1}{W \times H} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} |S(x,y) - GT(x,y)| \tag{19}$$

The MAE is the average pixel-wise absolute difference between the model saliency map $S(x,y)$ and the ground truth $GT(x,y)$. Utilizing this metric, the obtained value is lower and the performance better. It can be utilized to quantify the degree of approximation between the model saliency map $S(x,y)$ and the ground truth $GT(x,y)$.

4.2. Evaluation with Various Numbers of Superpixels

As mentioned previously, we adopted simple linear iterative clustering for generating compact and nearly uniform regions in the L*a*b* color space. To analyze the proposed method, we evaluated the effects of using various numbers of superpixels. We conducted comparative experiments on three datasets to evaluate whether using more superpixels affects performance.

As can be seen from Table 1, the performance of our method, in terms of MAE values, decreased with increasing numbers of superpixels for the MSRA and ECSSD datasets. In the DUT-OMRON dataset, the performance of our method, in terms of MAE values, increased with increasing numbers of superpixels. The MSRA 10K dataset contains many images with a single salient object, and the background structures of objects are primarily simple and smooth; therefore, utilizing a small number of superpixels can obtain more complete objects. The ECSSD dataset contains many images with more complex backgrounds. The DUT-OMRON dataset contains many images with relatively complex backgrounds and a single object; therefore, using a greater number of superpixels obtains more homogeneous objects.

The results are summarized in Table 1 in terms of precision, recall, F-measures, and MAE metrics. Based on the performance with various numbers of superpixels, we set the optimal number of superpixels as 20 in the proposed method.

TABLE 1 PERFORMANCE WITH VARIOUS NUMBERS OF SUPERPIXELS

Dataset	Metric	Number of superpixels							
		20	30	50	100	200	300	500	1000
ECSSD	P	0.730	0.725	0.730	0.732	0.734	0.735	0.735	0.732
	R	0.503	0.483	0.461	0.440	0.423	0.412	0.404	0.390
	F	0.607	0.600	0.594	0.589	0.584	0.579	0.575	0.565
	MAE	0.173	0.175	0.179	0.181	0.183	0.185	0.186	0.188
DUT-OMRON	P	0.527	0.538	0.536	0.531	0.532	0.531	0.531	0.531
	R	0.531	0.527	0.516	0.497	0.480	0.472	0.463	0.452
	F	0.480	0.488	0.484	0.477	0.471	0.469	0.466	0.463
	MAE	0.149	0.144	0.142	0.142	0.141	0.141	0.140	0.140

MSRA-10K	P	0.862	0.866	0.868	0.867	0.863	0.859	0.857	0.854
	R	0.603	0.583	0.554	0.526	0.505	0.493	0.483	0.470
	F	0.747	0.744	0.735	0.724	0.711	0.703	0.696	0.687
	MAE	0.118	0.124	0.130	0.137	0.142	0.145	0.1480	0.151

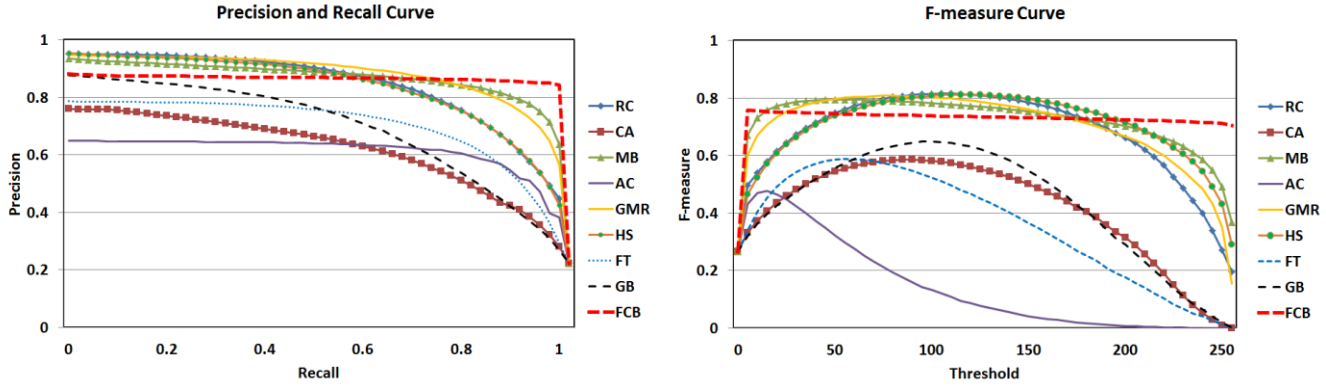
4.3 Evaluation with Various Benchmark Datasets

We selected the MSRA10K, ECSSD and DUT-OMRON datasets to evaluate the performance of our method. To avoid the figures becoming too crowded, only eight of the existing state-of-the-art methods were selected for comparison: CA [2], GB [8], HS [20], AC [3], FT [10], MB [17], GMR [18] and RC [29].

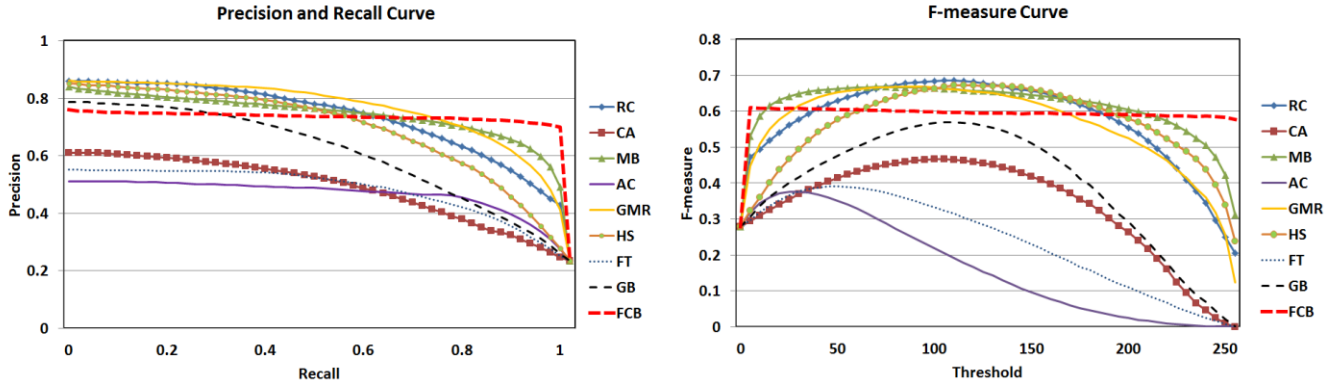
The source codes for MB [17] were run in the experiments, whereas for the other seven methods, we directly applied the saliency maps provided by the author of [11] at website [43].

The results of the quantitative evaluations are shown in Fig. 5. It consists of precision/recall and F-measure curves, and bar graphs for precision, recall, and F-measure utilizing adaptive thresholds.

The FCB method achieved good performance compared to several existing state-of-the-art methods in terms of the precision-recall (PR) curve and F-measure curve in Fig. 5. Compared to the MSRA10K, ECSSD and DUT-OMRON datasets, the FCB method achieved good precision scores that were only slightly lower than those of GMR and HS.



MSRA-10K



ECSSD

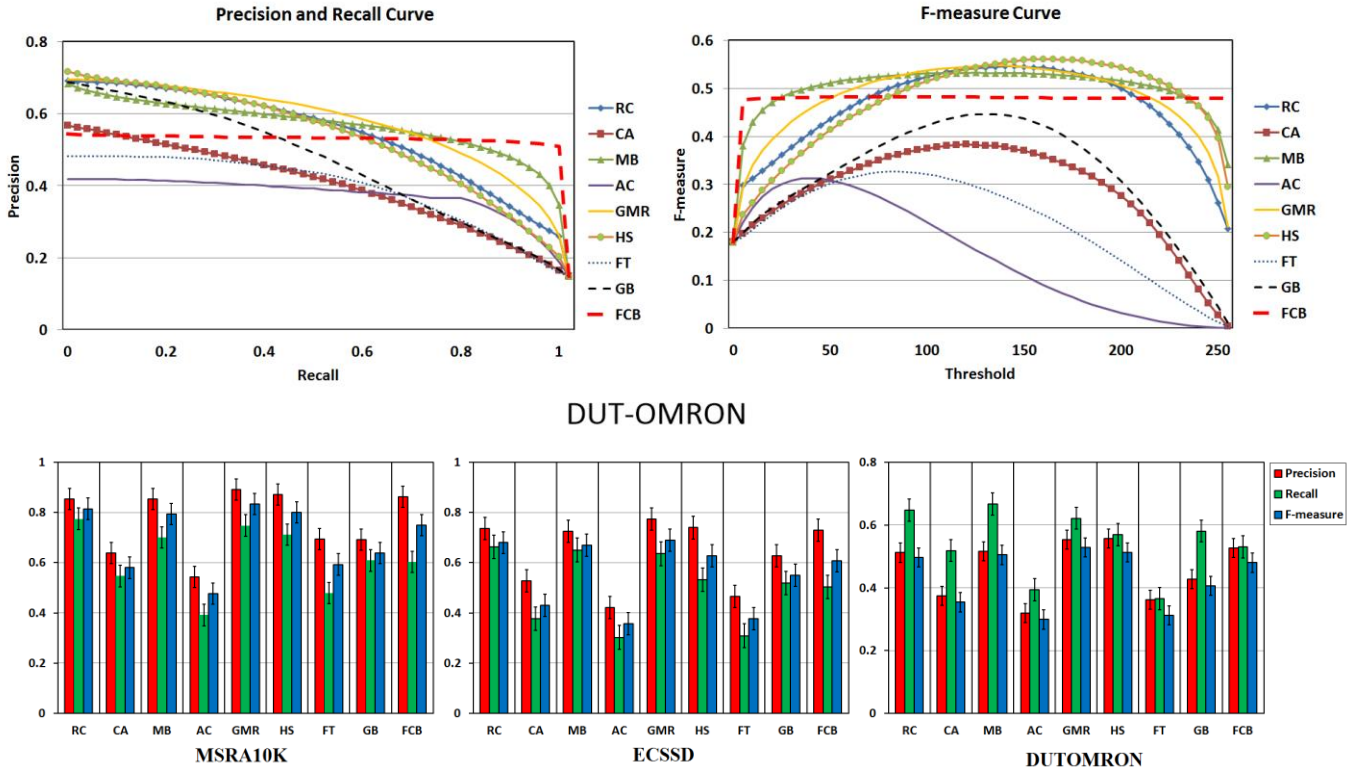
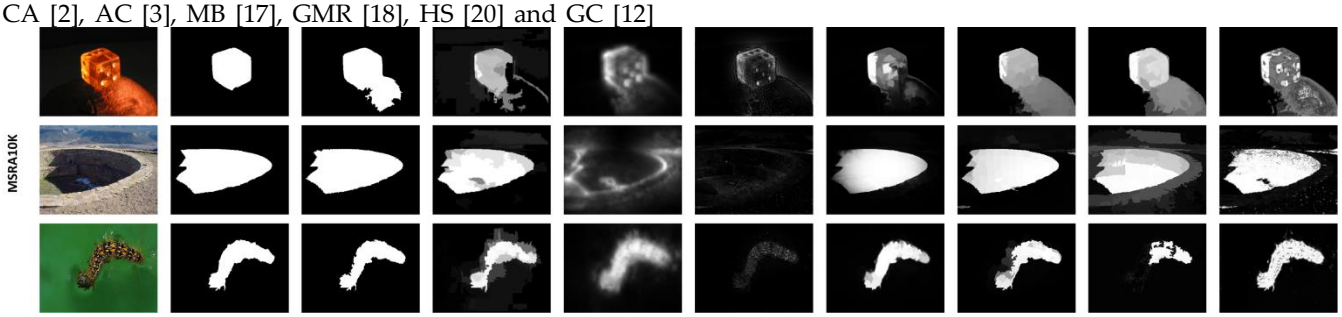


Fig. 5 Quantitative comparisons of various benchmark datasets (MSRA10K, ECSSD and DUT-OMRON) and the proposed FCB method. (Top) Precision/recall and F-measure curves, (bottom) precision, recall and F-measure performance evaluations using adaptive thresholds with error bars.

Some visual comparison examples of seven state-of-the-art methods are shown in Fig. 6. It is clear that our FCB inhibits background information, especially in datasets that include single objects or background objects that are primarily simple and smooth. The salient regions detected by our method were brighter than those of several existing state-of-the-art methods which show in Fig. 6.

The FCB method can detect salient objects robustly even when they touch the image boundaries, whereas other methods often fail in such cases. For instance, in the sixth row of images in Fig. 6, there is a woman lying on grass who is touching the image boundary. The RC [29], CA [2], AC [3], MB [17], GMR [18], HS [20] and GC [12] methods cannot handle this problem. In the ninth rows of images shows in Fig. 6, there are two sumo wrestlers whose legs are touching the image boundary. Again, the RC, CA, AC and MB methods cannot handle this problem. However, FCB method can easily reduce or address this problem.

It should be emphasized that the main aim of the present study was not to obtain the best precision scores, but rather, to achieve relatively high precision and recall simultaneously, and to detect salient objects robustly without sacrificing precision and recall even when salient objects touch the image boundaries.



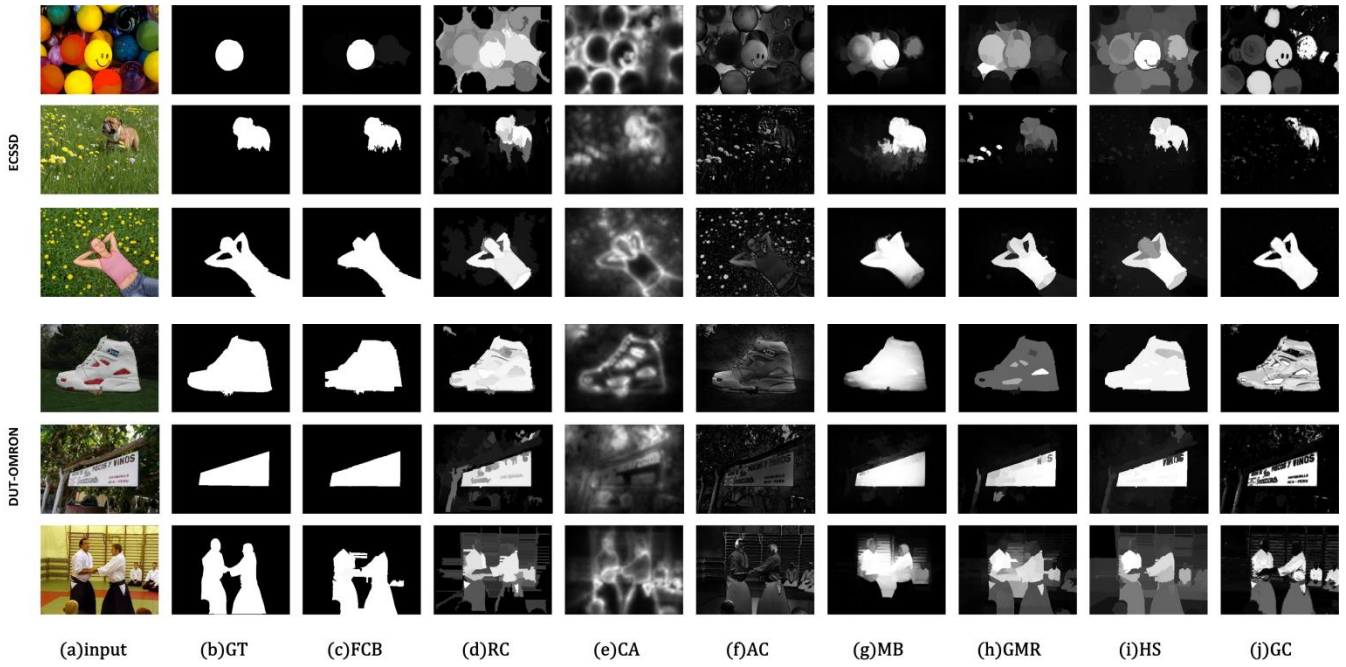


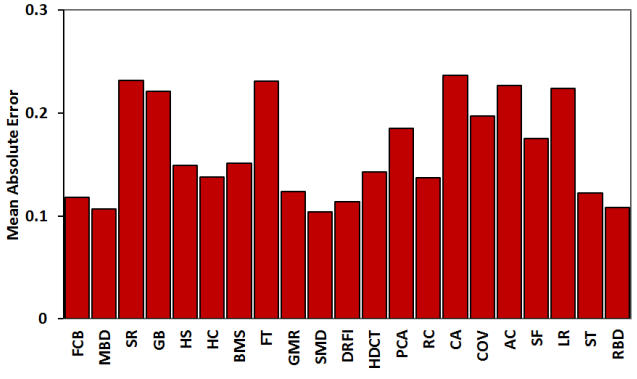
Fig. 6 visual comparison of saliency detection by our method (c) and seven other state-of-the-art methods on images from the MSRA10K, ECSSD, and DUT-OMRON benchmark datasets.

4.4 Evaluation by Mean Absolute Error (MAE)

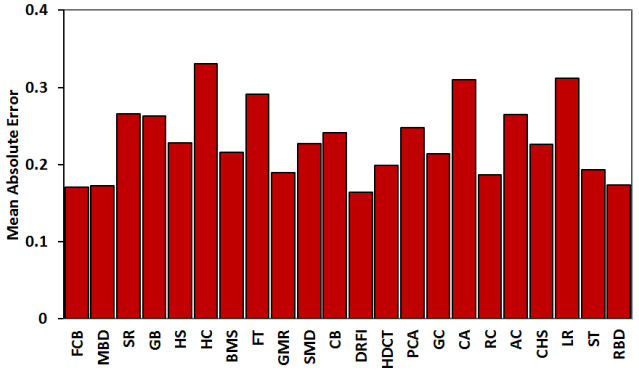
To determine the effects of the correct assignment of non-salient pixels and the importance of complete detection, similar performance comparisons were conducted using MAE metrics.

The FCB method was compared to MBD [17], SR [23], GB [8], HS [20], HC [11], BMS [31], FT [10], GMR [18], RC [29], CB [16], DRFI [49], HDCT [24], PCA [22], GC [12], CA [2], COV [50], AC [3], SF [30], RBD [15], LR [25] and ST [32].

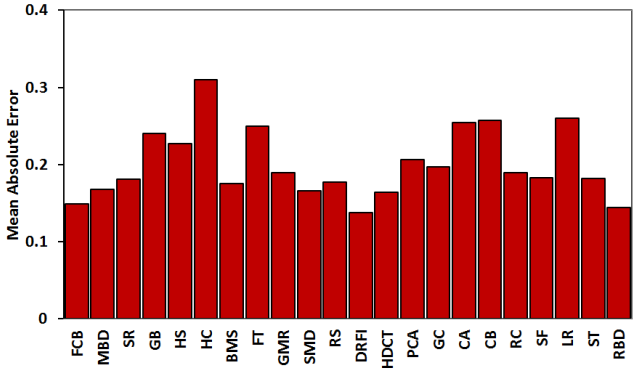
We directly utilized the evaluation data provided by Cheng [11], which can be download from website [43]. As for the RS [19], CHS [21] and SMD [26] methods, we adopted the original data that the authors reported in their papers using the default settings.



(a) MSRA10K



(b) ECSSD



(c) DUT-OMRON

Fig.7. Comparison of MAE results from various datasets and methods

The MAE results from various datasets are listed in Fig. 7. The FCB method achieved very good performance in

terms of MAE metrics with the MSRA10K, ECSSD and DUT-OMRON datasets.

The ECSSD dataset contains structurally complex natural images and multiple objects, and the FCB method performed second-best in terms of MAE metrics, only slightly lower than DRFT [49]. The DUT-OMRON dataset contains many images with relatively complex backgrounds and a single object, and again the FCB method achieved very good performance that was only slightly lower than DRFT [49] and RBD [15]. The experimental results validate that the proposed model can detect salient objects robustly in challenging images.

Based on the results of various benchmark datasets, FCB method demonstrated its ability to highlight salient objects by using regional color volume as the foreground and perceptually uniform color differences between regions. It can detect salient objects robustly, and precision and recall are not greatly sacrificed even when objects intersect the image boundaries.

4.5 Evaluation of Runtime

The experimental computations were run with an Intel Core™ i5-2400 3.10 GHz CPU and 16 GB RAM on a Windows 7 operating system. The FCB method was encoded with Visual C# 2013. In this section, some state-of-the-art methods were compared to the FCB method for the MSRA1000 dataset, including IT [1], GB [8], MBD [17], RC [29], FT [10], CB [16], HS [20], and GMR [18]. The average runtimes are shown in Fig. 8.

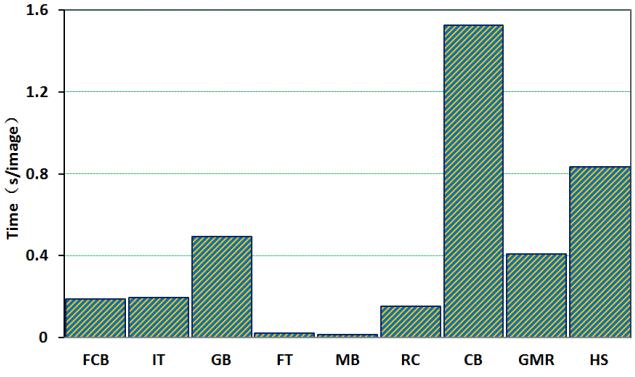


Fig. 8. Comparison of runtimes (seconds per image) on the MSRA-1000 dataset.

In the comparison of runtimes, the IT [1], GB [8], and CB [16] methods were coded in Matlab, FT [10] and RC [29] were coded in C++, GMR was coded in Matlab and C++, while the MB [17] and HS [20] methods directly utilized a Windows executable program written in C++.

Based on the comparison in Fig. 8, the FCB method demonstrates the advantage of a relatively low calculation burden. The computational cost mainly lies in the calculations of foreground saliency, center saliency, background saliency, and their normalizations.

4.6 Limitations

The FCB model has the advantage of a relatively low calculation burden using a small number of color superpixels. For example, with 20 color superpixels, it can detect salient objects robustly, even when they touch the image boundaries, without sacrificing precision and recall greatly. However, a limitation of FCB is that it is simultaneously sensitive to regional color volume and the perceptually uniform color differences between regions. It is difficult to perfectly balance these two effects.

Fig. 9. Some examples of failure with the FCB method

Figure 9 shows three typical examples of the limitations of our method. In some cases, salient regions are missing and some backgrounds are mistakenly considered to be salient objects.

5 SUMMARY

Foreground and background cues can help humans quickly understand a visual scene. In computer vision, however, it is difficult to detect salient objects when they touch the image boundary. Detecting salient objects robustly under such circumstances without sacrificing precision and recall can be challenging.

Color volume derived from the ellipsoid shape of the L*a*b* color space contains rich visual information, thus regional color volume together with the perceptually uniform color differences between regions was the basis of the present study. We proposed a novel saliency model to exploit color volume and perceptually uniform color differences for salient region detection, which involved foreground, center, and background cues.

Experimental results indicated that the proposed saliency model could detect salient objects robustly even when they touched the image boundary. It outperformed several state-of-the-art methods in terms of precision, recall, and F-measure. This was especially true when utilizing MAE metrics on the well-known benchmark datasets MSRA10K, ECSSD, and DUT-OMRON. Salient regions are brighter than those of several existing state-of-the-art methods

In further research, we plan to maintain the existing advantages of our method while improving its performance in terms of precision, recall, F-measure, and MAE

metrics via exploiting other properties within the L*a*b* color space.

REFERENCES

[1] L.Itti, C.Koch, 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.

[2] S. Goferman, L. Zelnik-Manor and A. Tal, "Context-aware saliency detection," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 2376-2383, 2010.

[3] R. Achanta, F. Estrada, P. Wils and S. Süsstrunk, "Salient Region Detection and Segmentation," International Conference on Computer Vision Systems, Springer Lecture Notes in Computer Science, pp. 66-75, 2008.

[4] G-H Liu, J-Y Yang, Z.Y. Li, "Content-based image retrieval using computational visual attention model, " Pattern Recognition,vol.48, no.8, pp. 2554-2566, 2015.

[5] Y.-F. Ma, H. Zhang, "Contrast-based image attention analysis by using fuzzy growing," Proceedings of ACM Multimedia, pp. 374-381, 2003.

[6] O.L. Meur, P.L. Callet, etc., "A coherent computational approach to model bottom-up visual attention, "IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.28, no.5, pp.802-817, 2006.

[7] A. Borji, L. Itti, "Exploiting local and global patch rarities for saliency detection," Proceedings of IEEE conference on computer vision and pattern recognition, pp. 478-485, 2012.

[8] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems, pp. 545–552, 2006.

[9] S. Frintrop, T. Werner, G. Garcia, "traditionla saliency reloaded: a good old model in new shape, " Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.82-90, 2015.

[10] R. Achanta, S. S. Hemami, F. J. Estrada and S. Süsstrunk , "Frequency-tuned salient region detection," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.1597-1604, 2009.

[11] M. M. Cheng, G. X. Zhang, N. J. Mitra, X. Huang and S. M. Hu, "Global contrast based salient region detection," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 409-416,2011.

[12] M-M Cheng, J. Warrell, W-Y Lin, S. Zheng, V.Vineet, N.Crook, "Efficient Salient Region Detection with Soft Image Abstraction," Proceedings of IEEE International Conference on Computer Vision, pp.1529 -1536, 2013.

[13] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," Proceedings of European Conference on Computer Vision, pp.29-42, 2012.

[14] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang, "Saliency detection via absorbing markov chain," Proceedings of IEEE International Conference on Computer Vision, pp. 1665–1672, 2013.

[15] W. Zhu, S.Liang, Y.Wei and J. Sun. "Saliency Optimization from robust background detection," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 2814 -2821, 2014.

[16] H.Jiang, J.Wang, Z.Yuan, T.Liu, N. Zheng, "Automatic salient object segmentation based on context and shape prior, " Proceedings of the British Machine Vision Conference, pp.110.1-110.12, 2011.

[17] J.Zhang, S.Sclaroff, Z. Lin, X.Shen, B.Price, R.Mech, "Minimum barrier salient object detection at 80 fps," Proceedings of IEEE International Conference on Computer Vision, pp.1404 -1412, 2015.

[18] C.Yang, L.Zhang, H.Lu, X.Ruan, M.Yang,"Saliency detection via graph-based manifold ranking," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 3166-3173, 2013.

[19] L. Zhang, C.Yang, H. Lu, X. Ruan, M-H. Yang. "Ranking Saliency," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.39, no.9, pp. 1892 -1904, 2017.

[20] Q. Yan, L. Xu, J. Shi and J. Jia, "Hierarchical saliency detection," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.1155-1162, 2013.

[21] J. Shi, Q. Yan, L. Xu, J. Jia. "Hierarchcal image saliency detection on extended CSSD. " IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.38, no.4, pp.717-729, 2016.

[22] R.Margolin, A.Tal, L.Zelnik-Manor, "What makes a patch distinct? " Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.1139 -1146, 2013.

[23] X. Hou, L. Zhang, "Saliency detection: A spectral residual approach," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition,pp.1–8, 2007.

[24] J Kim, D Han, YW Tai, J Kim, "Salient Region Detection via High-Dimensional Color Transform," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.883 -890, 2014.

[25] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 853 – 860, 2012.

[26] H. Peng; B. Li; H. Ling; W. Hu; W. Xiong; S. J. Maybank, "Salient Object Detection via Structured Matrix Decomposition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.39, no.4, pp. 818 – 832,2017.

[27] N. Li; J. Ye; Y. Ji; H. Ling; J. Yu, "Saliency detection on light field," IEEE Transactions on Pattern Analysis and Machine Intelligence , vol.39, no.8, pp. 1605 – 1616, 2017.

[28] J. Wang, H. Jiang, Z. Yuan, M-M Cheng, X. Hu, N. Zheng, "Salient Object Detection: A Discriminative Regional Feature Integration Approach," International Journal of Computer Vision, vol.123, no. 2, pp. 251–268, 2017.

[29] M-M. Cheng, M.J. Niloy, X. Huang, et al, " Global contrast based salient region detection, "IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.37, no.3, pp.569-582,2015.

[30] F. Perazzi, P. Krahenbuhl, Y. Pritch and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.733 – 740, 2012.

[31] J.Zhang, S.Sclaroff, "Saliency detection: a boolean map approach," Proceedings of IEEE International Conference on Computer Vision, pp.153-160, 2013.

[32] Z. Liu, W. Zou, O.L. Meur, "Saliency Tree: A Novel Saliency Detection Framework, " IEEE Transactions on Image Processing, vol.23, no.5, pp.1937 – 1952, 2014.

[33] S. He, R.W. Lau, W. Liu, Z. Huang, Q. Yang, "superCNN: A superpixelwise convolutional Neural Network for saliency object detection, " International Journal of Computer Vision, vol.115, no.3, pp.330–344, 2015.

[34] R.Zhao, W.Quyang, H.Li , X. Wang, "saliency detection by mutiscontext deep learning, " Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1265-1274,2015.

[35] T. Chen, L. Lin, L. Liu, X. Luo, X. Li, "DISC: Deep image saliency computing via progressive representation learning, " IEEE Transactions on Neural Networks and Learning Systems, vol. 27, no. 6, pp. 135-1149, 2016.

[36] X. Li, L. Zhao, L. Wei, M. Yang, F. Wu, Y. Zhuang, H. Ling, J.Wang, "Deepsaliency: Multi-task deep neural network model for salient object detection, " IEEE Transactions on Image Processing ,vol.25, no.8, pp.3919-3930, 2016.

[37] G. Li, Y. Yu, "Deep contrast learning for salient object detection, " Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.478- 487, 2016.

[38] N. Liu, J. Han, "DHSNet: Deep Hierarchical Saliency Network for Salient Object Detection," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 678 – 686, 2016.

[39] A. Toet, "Computational versus psychophysical bottom-up image saliency: A comparative evaluation study, "IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.33, no.11, pp.2131-2146, 2011.

[40] A. Borji, L. Itti, "State-of-the-art in visual attention modeling, "IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.35, no.1, pp.185-207, 2013

[41] G-H Liu, "Salient areas detection using color volume," Proceedings of IEEE Advanced Information Management,Communicates,Electronic and Automation Control Conference, pp. 474-478, 2016.

[42] 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.

[43] http://mmcheng.net/salobjbenchmark/

[44] W. Burger, M.J. Burge. Principles of Digital image processing: Core Algorithms, Springer, 2009.

[45] G-H. Liu, J-Y Yang," Content-based image retrieval using color difference histogram," Pattern Recognition, vol.46, no.1, pp188-198, 2013.

[46] http://sensing.konicaminolta.com.mx/2014/09/entendiendo-el-espacio-de-color-cie-lab/

[47] https://en.wikipedia.org/wiki/Lab_color_space

[48] A. Borji, M-M. Cheng, H. Jiang, J. Li, "Salient object detection: A benchmark, " IEEE Transactions on Image Processing,vol.24, no.12, pp.5706 – 5722,2015.

[49] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, S. Li, "Salient object detection: A discriminative regional feature integration approach", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2013.

[50] E.Erdem, A. Erdem, "Visual saliency estimation by nonlinearly integrating features using region covariances, " Journal of Vision, 13(4):11, 1-20, 2013.

[51] http://mmcheng.net/msra10k/

[52] http://www.cse.cuhk.edu.hk/leojia/projects/hsaliency/dataset.html

[53] http://saliencydetection.net/dut-omron/

Guang-Hai Liu received his Ph. D degree from the School of Computer Science and Technology, Nanjing University of Science and Technology (NUST). He is currently a professor with the College of Computer Science and Information Technology, Guangxi Normal University in China. His current research interests are in the areas of image processing, pattern recognition, and artificial intelligence.

Jing-Yu Yang received his B.Sc. degree in Computer Science from Nanjing University of Science and Technology (NUST), China. From 1982 to 1984 he was a visiting scientist at the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign. From 1993 to 1994 he was a visiting professor at the Department of Computer Science, Missourian University in 1998; he worked as a visiting professor at Concordia University in Canada. He is currently a professor and Chairman in the Department of Computer Science at NUST. He is the author of over 100 scientific papers in computer vision, pattern recognition, and artificial intelligence. He has won more than 20 provincial awards and national awards. His current research interests are in the areas of image processing, robot vision, pattern recognition, and artificial intelligence.