

Research Article

Inferring Visual Perceptual Object by Adaptive Fusion of Image Salient Features

Xin Xu,^{1,2} Nan Mu,^{1,2} and Hong Zhang^{1,2}

¹*School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430081, China*

²*Hubei Province Key Laboratory of Intelligent Information Processing and Real-Time Industrial System, Wuhan University of Science and Technology, Wuhan 430081, China*

Correspondence should be addressed to Xin Xu; xuxin0336@163.com

Received 19 September 2014; Revised 25 January 2015; Accepted 7 February 2015

Academic Editor: Marco Pérez-Cisneros

Copyright © 2015 Xin Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Saliency computational model with active environment perception can be useful for many applications including image retrieval, object recognition, and image segmentation. Previous work on bottom-up saliency computation typically relies on hand-crafted low-level image features. However, the adaptation of saliency computational model towards different kinds of scenes remains a challenge. For a low-level image feature, it can contribute greatly on some images but may be detrimental for saliency computation on other images. In this work, a novel data driven approach is proposed to adaptively select proper features for different kinds of images. This method exploits low-level features containing the most distinguishable salient information per image. Then the image saliency can be calculated based on the adaptive weight selection scheme. A large number of experiments are conducted on the MSRA database to compare the performance of the proposed method with the state-of-the-art saliency computational models.

1. Introduction

Saliency computational model with active environment perception can be useful for many applications including image retrieval, object recognition, and image segmentation. Generally, visual saliency can be defined as what captures human perceptual attention. Saliency detection plays an important role in image analysis and processing, which is capable of allocating the limited resources effectively. For example, the detecting of visual saliency can be effectively used to automatically zoom the “interesting” areas [1] or automatically crop the “important” areas in an image [2]. Object recognition algorithms can use the results of saliency detection to quickly locate the position of visual salient objects. Salient object detection can also reduce the interference of cluttered background to further improve the performance of image segmentation algorithm and image retrieval system [3].

Most of the existing saliency computational models are based on the bottom-up mechanism because the visual attention is generally driven by the low-level stimulus such as edge [4], color [5, 6], orientation [7], and symmetry [8].

These models typically contain two main procedures in saliency computation. The first step extracts low-level features from input image. Then the saliency map can be computed by fusing the extracted low-level features. In the past, low-level image features have been extensively studied. However, the selecting of proper features containing saliency information per-image is still complex and difficult to determine. The reason is mainly due to the lack of a well-defined feature which can exhaustively interpret saliency information in different images. Most of the existing saliency computational models face great difficulties in adaptively selecting low-level image features towards different images.

Aiming to address this problem, this paper puts forward an adaptive fusion scheme towards low-level image features for saliency detection. This method firstly extracts various low-level features to largely reflect saliency information of image. Then, the visual perceptual object can be detected by the adaptive weight selection towards these low-level image features. This method can retain the most significant low-level image features for saliency computation. The flowchart of the proposed method is illustrated in Figure 1.

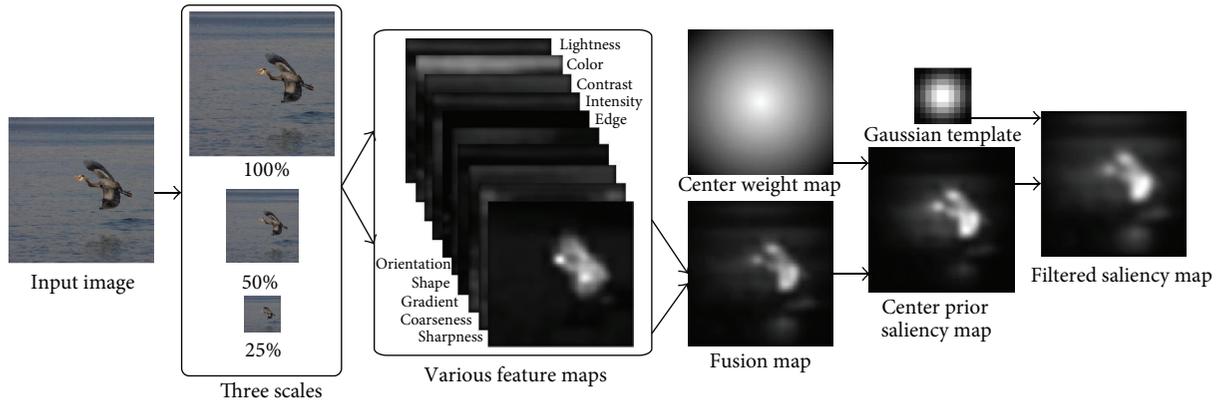


FIGURE 1: Flowchart of the proposed saliency computational model.

2. Related Works

Visual saliency reflects how much an image region or object stands out from its surrounding. Saliency computational model aims to provide a numerical measure indicating the degree of visual saliency, and is useful for a wide range of applications. Consequently, a number of computational approaches for salient object detection have been developed in recent years. Based on the biologically plausible visual attention architecture [9] and the feature integration theory [10], Itti et al. [11] proposed the bottom-up saliency model which depends only on low-level image features. The salient object in image can be determined using dynamic neural networks for feature fusion. Following this model, many state-of-the-art saliency computational models focus on low-level features, such as color and contrast [12, 13].

For example, Achanta et al. [14, 15] used the luminance and color features to detect salient object in image. This method calculated the contrast between local image region and its surrounding region and utilized the average color vector difference to obtain the saliency value of image. Aziz and Mertsching [8], Liu et al. [16], and Cheng et al. [17] proposed a contrast computational model based on region segmentation. These methods took use of the image segmentation algorithm to divide the image into different regions according to the homogeneity of different low-level image features such as luminance, texture, and color.

Most of the existing saliency computational models are based on low-level image features. Consequently, there are various formulations for exploiting well-defined salient features. Gao and Vasconcelos [18] studied the statistics of nature images and constructed an optimal detector based on the discriminate saliency. This method can effectively integrate the saliency features. Klein and Frinrop [19] detected the salient object by reconstituting the cognitive visual attention model. Their approach computed the saliency of all feature channels in an information-theoretic way. The optimal features can be determined by using Kullback-Leibler divergence (KLD) to fuse different features. Lu et al. [20] put forward the diffusion-based salient object detection model. This model can learn the optimal saliency seeds by combining two kinds

of features: the bottom-up saliency maps and midlevel vision cues.

However, most of the existing saliency computational models fail to consider the adaptability of different low-level features towards different images. Some low-level features which contribute greatly on some images may actually be detrimental to the detection of saliency on other images. In this paper, a novel data driven approach is proposed to adaptively select proper features for different images. This method exploits the most distinguishable low-level features containing salient information per image. Then the image saliency can be calculated based on the adaptive weight selection scheme.

The rest of this paper is organized as follows. Section 3 details the extraction of different low-level image features. Section 4 puts forward the adaptive feature fusion scheme. The experimental results are given in Section 5. And finally, the conclusions are drawn in Section 6.

3. Low-Level Image Feature Extraction

According to physiological experiments, it can be found that human attention towards image is mainly driven by low-level image features [21]. In order to fully and accurately describe the saliency information in image, ten low-level image features are chosen according to their visual properties. These ten low-level features are lightness, color, contrast, intensity, edge, orientation, shape, gradient, coarseness, and sharpness. In this section, we describe the extraction of these features and analyze the contributions of each feature in saliency computation.

The low-level features are extracted using the block-based method from input image (denoted by I), which is divided into 8×8 blocks (denoted by b) with 50% overlap. Let $F_n(I)$, $n = 1, \dots, 10$, denote the saliency map corresponding to each low-level feature; then $F_n(b)$ can be used to denote the saliency map of block b .

3.1. Lightness Feature Extraction. Lightness is the perception attribute of human vision system for visible objects radiation

or glow amount. The lightness feature may affect the performance of other features to a certain extent.

The lightness feature measures how the brightness of each block is different from the average brightness of the image. Therefore, the object having a high brightness can be treated as a salient object by the human eye gaze. Let $F_1(b)$ denote the Euclidean distance of the average lightness value between each image block b and the image I . The larger distance stands for the greater lightness value difference. Specific methods are shown as follows.

Firstly, the input image is converted from the RGB color space to a uniform LAB color space. Then, the lightness feature $F_1(b)$ can be obtained by

$$F_1(b) = |\bar{L}(b) - \bar{L}(I)|, \quad (1)$$

where \bar{L} represents the average lightness component L in LAB color space. The lightness feature can well distinguish the illumination differences of object in image.

3.2. Color Feature Extraction. As the salient object has a strong contrast or a strong variation, the color value of the image object is relatively far from the average color value of the whole image. The background region can be seen as the smooth area in which the color value is closed to the average color value. Therefore, the color feature can be extracted by calculating the Euclidean distance of the average color value between each image block and the background region. The color feature $F_2(b)$ of the image block b can be computed via

$$F_2(b) = \sqrt{(\bar{A}(b) - \bar{A}(I))^2 + (\bar{B}(b) - \bar{B}(I))^2}, \quad (2)$$

where \bar{A} and \bar{B} represent the average A and B color components in LAB color space, respectively.

Most of the existing saliency computational models are conducted in the RGB or LAB color feature space. RGB is mainly used for the color representation of the image, while CIELab provides a representation of color that corresponds to how observers perceive chromatic differences. Thus, this method extracts the color feature in the LAB color space.

Color feature can simply describe the global distribution of colors in an image. Thus, it can measure the proportion of different colors accounting for the whole image. The color feature is especially suitable for the image which does not take the spatial position of the object into account.

3.3. Contrast Feature Extraction. The saliency of the image block b in image I depends on the difference between the image block and its surrounding, while the visual characteristic of the image block itself does not determine its saliency. Thus, the greater the difference between the image block and its surrounding, the more likely it is a salient region. The calculation method to extract the contrast feature $F_3(b)$ of the image block b is shown as follows.

First, convert each block to luminance image block via

$$l(b) = (i + jb)^\varepsilon, \quad (3)$$

where $i = 0.7656$, $j = 0.0364$, and $\varepsilon = 2.2$ denote the display conditions of the Adobe RGB color space. The contrast feature of the image block b can be expressed as

$$F_3(b) = \begin{cases} \frac{\sigma(b)}{\mu(b)}, & \mu(b) > 0, \\ 0, & \mu(b) = 0, \end{cases} \quad (4)$$

where $\sigma(b)$ and $\mu(b)$ represent the standard deviation and the average value of $l(b)$, respectively. The contrast is critically important on the visual attention. Generally, the greater the contrast is, the sharper the image can be.

3.4. Intensity Feature Extraction. The intensity feature (denoted by $S(I)$) can also be treated as a lightness feature and can be obtained by averaging the three RGB color channels (denoted by $R(I)$, $G(I)$, and $B(I)$). Besides, the intensity feature can also reflect the brightness information and the color variations of image, which can represent human subjective feeling. The intensity feature $F_4(b)$ of image block b can be generated by calculating the Euclidean distance between the intensity value of the image block $S(b)$ and the whole image $S(I)$:

$$F_4(b) = |\bar{S}(b) - \bar{S}(I)|, \quad (5)$$

where $\bar{S}(b)$ and $\bar{S}(I)$ represent the average intensity values of $S(b)$ and $S(I)$, respectively. The human eye perception towards the brightness is mainly associated with the luminous intensity of observed objects.

3.5. Edge Feature Extraction. The edge refers to a collection of image pixel in which the grayscale intensity has a strong contrast change. Thus, we use the edge feature to capture the image region with dramatic brightness variations. Let $E(b)$ denote the binary map obtained through the Roberts edge detection. The edge feature $F_5(b)$ of the image block b is then achieved by computing the average value of $E(b)$ as

$$F_5(b) = \mu(E(b)). \quad (6)$$

The edge in image is often associated with the discontinuity of image grayscale.

3.6. Orientation Feature Extraction. Orientation feature is similar to the color feature, which is also an integrity feature. Based on the Itti model, the proposed method calculates the orientation feature by running the Gabor filter on grayscale image (denoted by $g(x, y)$) of input image I . The Gabor wavelet can extract the direction feature effectively and can eliminate the redundant information at the same time. The two-dimensional Gabor function $G(x, y)$ is shown as follows:

$$G_{\lambda, \theta, \varphi, \sigma, \gamma}(x, y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \varphi\right),$$

$$x' = x \cos \theta + y \sin \theta,$$

$$y' = -x \sin \theta + y \cos \theta, \quad (7)$$

where $\gamma = 0.5$ represents the spatial aspect ratio and $\sigma = 0.56\lambda$ represents the standard deviation of the Gaussian factor. The parameters λ , φ , and θ denote the wavelength, the phase offset, and the angle, respectively.

The 2D Gabor operator based orientation image feature extraction (denoted by $O_\theta(I)$, $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$) can be obtained using the following form of convolution:

$$O_\theta(I) = I(x, y) * G_\theta(x, y). \quad (8)$$

Let $F_6(b)$ denote the orientation feature of image block b , which can be obtained by computing the Euclidean distance between the orientation image block $O_\theta(b)$ and the orientation image $O_\theta(I)$ via

$$F_6(b) = \sum_{\theta=\{0^\circ, 45^\circ, 90^\circ, 135^\circ\}} |\overline{O}_\theta(b) - \overline{O}_\theta(I)|, \quad (9)$$

where $\overline{O}(b)$ and $\overline{O}(I)$ represent the average orientation values of $O(b)$ and $O(I)$, respectively. The orientation feature has global properties and thus can enable the saliency information with good feasibility and stability even in complex scenes.

3.7. Shape Feature Extraction. The shape feature can be extracted using Hu's moment invariants, which has the invariant properties of rotation, translation, and scale. The Hu moment invariants are commonly used to identify the large objects in an image and can better describe the shape feature of an object. The proposed method calculates the 7 moments (denoted by M_i , $i = 1, \dots, 7$) by Hu's method, which is obtained by normalizing the central moments through orders two and three.

Let $M_i(b)$ and $M_i(I)$, $i = 1, \dots, 7$, denote the 7 moments of the image block b and the input image I , respectively. Thus, the shape feature (denoted by $F_7(b)$) of b can be expressed as the Euclidean distance between $M_i(b)$ and $M_i(I)$:

$$F_7(b) = \sum_{i=\{1,2,\dots,7\}} |M_i(b) - M_i(I)|. \quad (10)$$

The shape feature is not sensitive to the lightness and the contrast changes and thus can effectively reduce the influence of lightness.

3.8. Gradient Feature Extraction. The gradient feature is sensitive to the gradient variation; however, it is not sensitive to the grayscale of the image. The image gradient can not only be able to capture the contour, silhouette, and some texture information, but also further weaken the influence of illumination. Let $g(x, y)$ be the grayscale at pixel (x, y) in the image region, with the size $M \times N$; the gradient feature (denoted by $F_8(b)$) of image block b can be calculated by averaging the abscissa squared gradient (denoted by

$g_x(b)$) and the ordinate squared gradient (denoted by $g_y(b)$) through

$$\begin{aligned} g_x(b) &= \sum_{x=0}^{M-2} \sum_{y=0}^{N-1} (g(x+1, y) - g(x, y))^2, \\ g_y(b) &= \sum_{x=0}^{M-1} \sum_{y=0}^{N-2} (g(x, y+1) - g(x, y))^2, \\ F_8(b) &= \mu \left(\frac{g_x(b) + g_y(b)}{2} \right). \end{aligned} \quad (11)$$

Gradient value can describe the magnitude of the dramatic changes of the pixel values. Thus, the gradient map constituted by the pixel gradient values can reflect the local grayscale changes in image.

3.9. Coarseness Feature Extraction. Coarseness is the fundamental perceptual texture feature. It can measure the particle size of the texture pattern. The larger particle size means the coarser image texture. The coarseness feature can be calculated as follows.

Firstly, the average gray value (denoted by $A_k(x, y)$, $i = 1, 2, \dots, 5$) of the neighborhood with size $2^k \times 2^k$ in image I is calculated as

$$A_k(x, y) = \frac{\sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y-2^{k-1}-1} g(x, y)}{2^{2k}}, \quad (12)$$

where $g(x, y)$ is the gray value at pixel (x, y) in the active window.

Then, for each pixel, the average intensity difference (denoted by $E_{k,h}(x, y)$ and $E_{k,v}(x, y)$) can be calculated between the nonoverlapping neighborhoods in the horizontal and vertical directions, respectively:

$$\begin{aligned} E_{k,h}(x, y) &= |A_k(x + 2^{k-1}, y) - A_k(x - 2^{k-1}, y)|, \\ E_{k,v}(x, y) &= |A_k(x, y + 2^{k-1}) - A_k(x, y - 2^{k-1})|. \end{aligned} \quad (13)$$

Finally, the optimal size can be set by k which gives the highest value of E . The coarseness (denoted by C) is the average of $S_{\text{best}}(x, y) = 2^k$:

$$C = \frac{1}{m \times n} \sum_{x=1}^m \sum_{y=1}^n S_{\text{best}}(x, y). \quad (14)$$

Let $C(b)$ and $C(I)$ denote the coarseness of the image block b and image I , respectively. Thus, the coarseness feature (denoted by $F_9(I)$) of b can be expressed as the Euclidean distance between $C(b)$ and $C(I)$:

$$F_9(b) = |C(b) - C(I)|. \quad (15)$$

Coarseness feature represents the surface properties of the whole image and can well describe the integrity of the salient object. Meanwhile, the coarseness feature has good rotation invariance; it can effectively resist the interference of noise.

3.10. Sharpness Feature Extraction. The sharpness feature measures how the acutance of each region is different from its surrounding, which can indicate the contrast of the adjacent region. The proposed method extracts the sharpness value $\phi(p)$ at position p by computing the convolution between the input image I and the first-order derivatives of the Gaussian via

$$\phi(p) = \sum_x \sum_y [g(x, y) * G_x^\sigma(x, y)]^2 + [g(x, y) * G_y^\sigma(x, y)]^2, \quad (16)$$

where $g(x, y)$ is the grayscale at pixel (x, y) in the image region. $G_x^\sigma(x, y)$ and $G_y^\sigma(x, y)$ represent the first-order derivatives of the Gaussian in the vertical and the horizontal directions, respectively. σ is the scale of the Gaussian filter.

The Gaussian derivative method can measure the acutance variation and suppress the influence of noise and illumination. Let $\bar{\phi}(b)$ and $\bar{\phi}(I)$ denote the average sharpness value of the image block b and the input image I , respectively. Thus, the sharpness feature (denoted by $F_{10}(b)$) of b can be expressed as the Euclidean distance between $\bar{\phi}(b)$ and $\bar{\phi}(I)$:

$$F_{10}(b) = |\bar{\phi}(b) - \bar{\phi}(I)|. \quad (17)$$

Sharpness represents the image definition and the edge acuteness. The higher the sharpness is, the higher the image contrast can be. The sharpness feature can be less susceptible to local variations.

4. Adaptive Fusion of Low-Level Image Features

According to the discrete degree and clarity of the saliency map of each low-level feature, different weights are assigned to different features. Thus, low-level features that contribute greatly on some images will assign a large weight; other low-level features that may be detrimental to the detection of saliency will assign a low weight or completely ignore during the saliency computation. Let ν_n denote the statistical validity and ω_n denote the weights of different feature maps $F_n(I)$, $n = 1, 2, \dots, 10$. The ν_n can be defined as

$$\nu_n = \sigma n^2 + \kappa n, \quad (18)$$

where σn^2 and κn represent the variance and the kurtosis of $F_n(I)$, respectively.

The weights ω_n of different feature maps $F_n(I)$ are determined by the numerical magnitudes of the statistical validity ν_n via

$$\omega_n = \begin{cases} 1, & \text{if } \nu_n = \nu_1^* \\ \frac{4}{5}, & \text{if } \nu_n = \nu_2^* \\ \frac{3}{5}, & \text{if } \nu_n = \nu_3^* \\ \frac{2}{5}, & \text{if } \nu_n = \nu_4^* \\ \frac{1}{5}, & \text{if } \nu_n = \nu_5^* \\ 0, & \text{otherwise,} \end{cases} \quad (19)$$

where $\nu_i^* = \text{sort}\{\nu_i\}$ using the descending order. The proposed method assigns different weights according to numerical sort of ν_n .

The final fusion map (denoted by F_I) is calculated by the weighted sum of the ten feature maps $F_n(I)$, $n = 1, 2, \dots, 10$:

$$F_I = \frac{\sum_n \omega_n F_n(I)}{\sum_n \omega_n}. \quad (20)$$

To enhance the robustness of detection process and achieve a preferable visual effect, the proposed method is performed in three scales {100%, 50%, 25%} to better restrain the background information.

The obtained saliency map is then refined using center prior principle to enhance visual effect. When humans watch a picture, they will naturally gaze on the objects next to the center of image [22]. Thus, in order to obtain the saliency objects closer to the human visual fixations, more weight is needed to add to the center of image. Therefore, a feature (denoted by $f_c(b)$) is included to indicate the distance between each image block and the center of image. For each image block feature $F_n(b)$, $n = 1, \dots, 10$, it can be recalculated via

$$F_n^*(b) = F_n(b) f_c(b),$$

$$f_c(b) = 1 - \frac{\sqrt{(r(b) - C^*)^2 + (c(b) - R^*)^2}}{\sqrt{(M/2)^2 + (N/2)^2}}, \quad (21)$$

where $(r(b), c(b))$ denote the upper-left coordinate of image block b , (C^*, R^*) represent the center coordinate of the estimated image region, and M and N denote the width and height of the image region, respectively.

The feature map $F_n(I)$ of the whole image I can be generated by combining the image feature $F_n^*(b)$ of all the blocks. The normalized feature map is calculated by

$$F_n(I) = \frac{F_n(I) - \min(F_n(I))}{\max(F_n(I)) - \min(F_n(I))}. \quad (22)$$

Finally, the generated saliency map is smoothed by a Gaussian filter (the template size is 10×10 , and σ is 2.5).

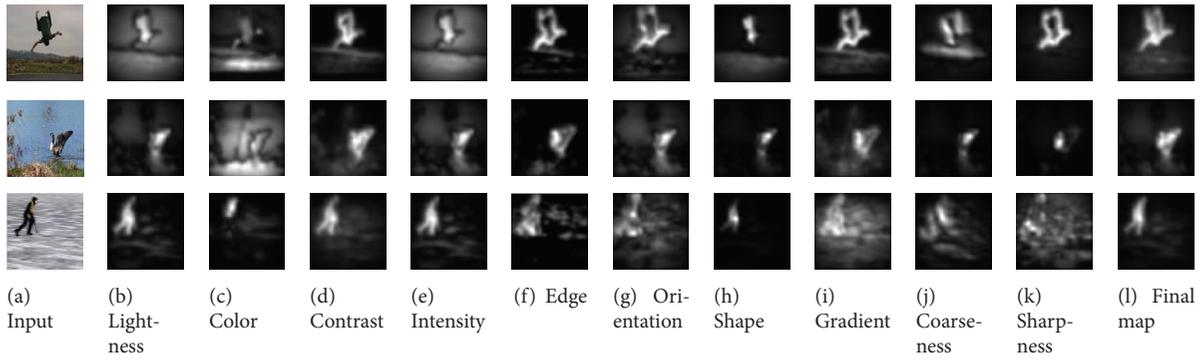


FIGURE 2: Saliency maps of each low-level features. (a) Input image, ((b)–(k)) corresponding saliency maps of low-level image features respectively, (l) the final saliency map.

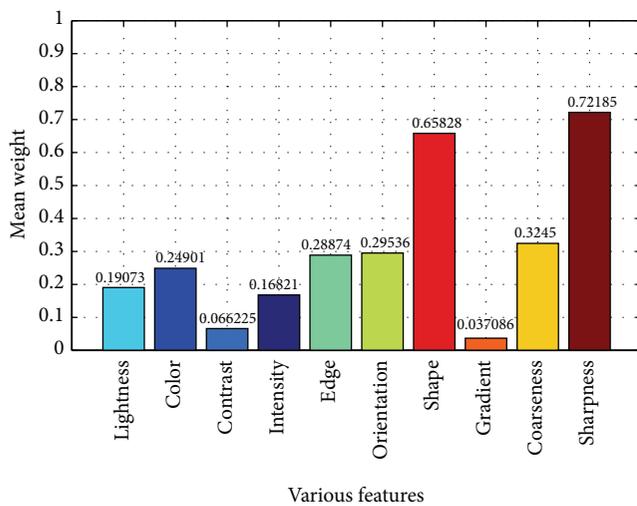


FIGURE 3: The average weight distribution of the ten low-level image features.

The saliency maps of each low-level feature and the final saliency maps are shown in Figure 2.

As illustrated in Figure 2, the ten low-level image features have their own advantages and disadvantages towards different images. On the contrary, the final saliency map can adaptively fuse the optimal features to achieve better performance.

In addition, we also provide the statistical of the ten low-level image features in different images. The result from different images is shown in Figure 3. As illustrated in Figure 3, the sharpness feature and the shape feature have good stability in our testing. Their weights have little change due to the deviation caused by local contrast transformation. On the contrary, the gradient feature can only reflect the local differences.

5. Experimental Results

The performance evaluation is conducted using the MSRA salient object database [16]. This database contains over

20 000 images, which includes two parts: (i) image set A, containing 20,000 images, and the principle salient objects are labeled by three users, and (ii) image set B, containing 5,000 images, and the principle salient objects are labeled by nine users. The proposed multifeature fusion (MF) approach is compared with the other seven state-of-the-art methods: *Itti's* (IT) method [11], *spectral residual* (SR) method [23], *saliency using natural statistics* (SUN) method [24], *frequency-tuned* (FT) method [25], *S3* method [26], *nonparametric* (NP) method [27], and *context-aware* (CA) method [28].

Figure 4 illustrates the performance comparison of these various salient region detection methods. As can be seen from Figure 4, the saliency maps extracted using our proposed method are more consistent with the ground-truth rectangle, and the detected saliency objects are more similar to the ground-truth binary masks. The approaches developed in [23, 24, 26] fail to detect and clearly identify the location of salient object from complex background. The saliency maps generated using methods [11, 27] look rather blurry and are difficult to clearly distinguish the salient region. The saliency map of method [25] retains a lot of background information. The CA method [28] can achieve good detection performance on some images; however this method is unable to highlight the entire salient object itself.

The objective assessment is implemented by computing the *true positive rate* (TPR) and the *false positive rate* (FPR). Given the ground-truth binary masks $G(x, y)$ and the obtained saliency map $F_t(x, y)$ ($0 \leq F_t(x, y) \leq 1$), a threshold t ($0 \leq t \leq 1$) is used to obtain the binary masks $B_t(x, y)$, in which 0 denotes the background and 1 denotes the salient objects. The TPR and FPR can be computed via

$$\begin{aligned} \text{TPR} &= E \left(\prod_t B_t(x, y) \cdot G(x, y) \right), \\ \text{FPR} &= E \left(\prod_t (1 - B_t(x, y)) \cdot G(x, y) \right). \end{aligned} \quad (23)$$

Figure 5 shows the TPR and FPR results of the seven methods and the proposed method. As seen in Figure 5, the overall performance of the proposed method is more excellent than the other seven methods.



FIGURE 4: Saliency maps obtained from the proposed method (MF) and state-of-the-art saliency computational models.

TABLE 1: The precision, recall, and F -measure of the saliency computational models.

	IT [11]	SR [23]	SUN [24]	FT [25]	S3 [26]	NP [27]	CA [28]	MF
Precision	0.3548	0.5662	0.2769	0.6845	0.2159	0.2132	0.4894	0.7155
Recall	0.9477	0.4177	0.7491	0.6519	0.6991	0.9522	0.8770	0.8150
F -measure	0.4391	0.4951	0.3291	0.6443	0.2715	0.2781	0.5580	0.7370

TABLE 2: Average execution time (in seconds) for the saliency computational models.

	IT [11]	SR [23]	SUN [24]	FT [25]	S3 [26]	NP [27]	CA [28]	MF
Run time	1.0047	0.1810	1.7783	0.1710	8.4211	1.4895	68.3333	10.3720

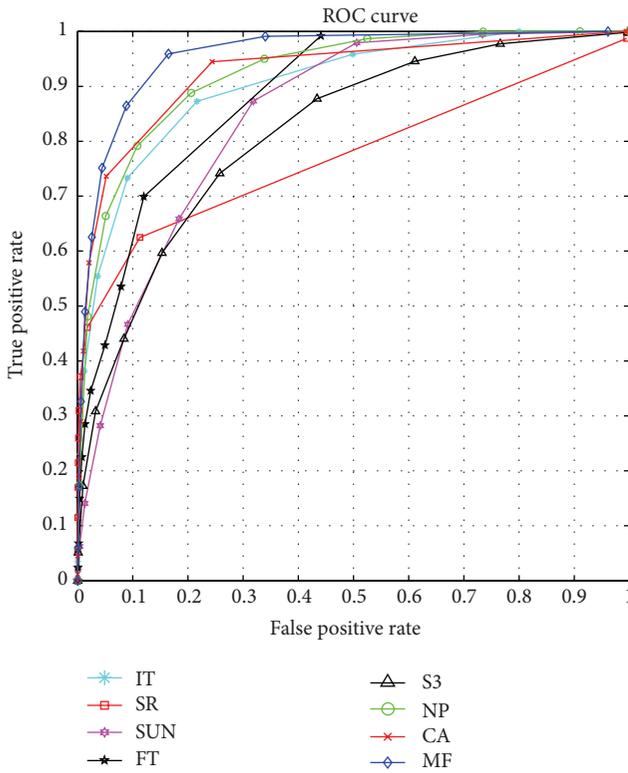


FIGURE 5: ROC curves obtained from the proposed method and state-of-the-art saliency computational models.

Given the generated saliency map $F_I(x, y)$, we set a threshold T (computed by Otsu's method) to segment the saliency objects. The binary mask is denoted by $B_I(x, y)$.

Let G and B_I denote the ground-truth binary masks and binary masks of the proposed approach, respectively. The Precision = $R(B_I \cap G)/R(B_I)$ and Recall = $R(B_I \cap G)/R(G)$, where $R(\cdot)$ represent the salient region. The evaluation criterion F -measure can be computed via:

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

The proposed method uses $\beta^2 = 0.5$ to weigh the precision and recall. The precision, recall, and F -measure of these methods are shown in Table 1.

Finally, we compare the computational complexity of the different saliency computational models discussed. These models are implemented using the MATLAB programming language and run on a PC with a Pentium G2020 CPU and a 4 GB RAM. Table 2 shows the results of the proposed method and the other methods. The proposed method yields slightly higher computational load than the conventional approach; however, the proposed method can achieve more accurate saliency detection in various images.

6. Conclusion

In this paper, a novel feature selection scheme is proposed to adaptively select the proper features for different images. This method exploits the most distinguishable salient information in ten low-level features per image. The generated saliency map can highlight the salient object in different images even containing complex background. A large number of experiments are conducted on the MSRA database to compare the performance of the proposed method with the state-of-the-art saliency computational models. And the experimental results indicate that the proposed method outperforms the state-of-the-art saliency computational models to achieve better performance.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (nos. 61440016 and 61373109), the Natural Science Foundation of Hubei Provincial of China (no. 2014CFB247), and the Project of the Key Laboratory for Metallurgical Equipment and Control of Ministry of Education (no. 2013B08).

References

- [1] L.-Q. Chen, X. Xie, X. Fan, W.-Y. Ma, H.-J. Zhang, and H.-Q. Zhou, "A visual attention model for adapting images on small displays," *Multimedia Systems*, vol. 9, no. 4, pp. 353–364, 2003.

- [2] F. Stentiford, "Attention based auto image cropping," in *Proceedings of the 5th International Conference on Computer Vision Systems*, Bielefeld, Germany, March 2007.
- [3] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Random walks on graphs to model saliency in images," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '09)*, pp. 1698–1705, June 2009.
- [4] Q. Deng and Y. Luo, "Edge-based method for detecting salient objects," *Optical Engineering*, vol. 50, no. 5, Article ID 057007, 2011.
- [5] K. Fu, C. Gong, J. Yang, Y. Zhou, and I. Yu-Hua Gu, "Superpixel based color contrast and color distribution driven salient object detection," *Signal Processing: Image Communication*, vol. 28, no. 10, pp. 1448–1463, 2013.
- [6] J. Kim, D. Han, Y.-W. Tai, and J. Kim, "Salient region detection via high-dimensional color transform," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 883–890, Columbus, Ohio, USA, June 2014.
- [7] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Salient region detection by modeling distributions of color and orientation," *IEEE Transactions on Multimedia*, vol. 11, no. 5, pp. 892–905, 2009.
- [8] M. Z. Aziz and B. Mertsching, "Fast and robust generation of feature maps for region-based visual attention," *IEEE Transactions on Image Processing*, vol. 17, no. 5, pp. 633–644, 2008.
- [9] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, no. 4, pp. 219–227, 1985.
- [10] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, no. 1, pp. 97–136, 1980.
- [11] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [12] Y. Lu, W. Zhang, H. Lu, and X. Xue, "Salient object detection using concavity context," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 233–240, November 2011.
- [13] K. Shi, K. Wang, J. Lu, and L. Lin, "PISA: pixelwise image saliency by aggregating complementary appearance contrast measures with spatial priors," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 2115–2122, June 2013.
- [14] R. Achanta, F. Estrada, P. Wils, and S. Susstrunk, "Salient region detection and segmentation," in *Computer Vision Systems: 6th International Conference, ICVS 2008 Santorini, Greece, May 12–15, 2008 Proceedings*, vol. 5008 of *Lecture Notes in Computer Science*, pp. 66–75, Springer, Berlin, Germany, 2008.
- [15] R. Achanta and S. Süssstrunk, "Saliency detection using maximum symmetric surround," in *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP '10)*, pp. 2653–2656, September 2010.
- [16] T. Liu, J. Sun, N.-N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, June 2007.
- [17] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 409–416, June 2011.
- [18] D. Gao and N. Vasconcelos, "Bottom-up saliency is a discriminant process," in *Proceedings of the IEEE 11th International Conference on Computer Vision (ICCV '07)*, pp. 1–6, Rio de Janeiro, Brazil, October 2007.
- [19] D. A. Klein and S. Frintrap, "Center-surround divergence of feature statistics for salient object detection," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 2214–2219, November 2011.
- [20] S. Lu, V. Mahadevan, and N. Vasconcelos, "Learning optimal seeds for diffusion-based salient object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14)*, pp. 2790–2797, IEEE, Columbus, Ohio, USA, June 2014.
- [21] C. T. Vu and D. M. Chandler, "Main subject detection via adaptive feature selection," in *Proceedings of the 16th IEEE International Conference on Image Processing (ICIP '09)*, pp. 3101–3104, Cairo, Egypt, November 2009.
- [22] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *Proceedings of the 12th International Conference on Computer Vision (ICCV '09)*, pp. 2106–2113, September–October 2009.
- [23] X. Hou and L. Zhang, "Saliency detection: a spectral residual approach," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, June 2007.
- [24] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell, "SUN: a Bayesian framework for saliency using natural statistics," *Journal of Vision*, vol. 8, no. 7, article 32, 2008.
- [25] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 1597–1604, Miami, Fla, USA, June 2009.
- [26] C. T. Vu, T. D. Phan, and D. M. Chandler, " S_3 : a spectral and spatial measure of local perceived sharpness in natural images," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 934–945, 2012.
- [27] N. Murray, M. Vanrell, X. Otazu, and C. A. Parraga, "Saliency estimation using a non-parametric low-level vision model," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 433–440, June 2011.
- [28] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1915–1926, 2012.



Hindawi

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

