Research Article

An Improved Saliency Detection Approach for Flying Apsaras in the Dunhuang Grotto Murals, China

Zhong Chen,[1] Shengwu Xiong,[1] Qingzhou Mao,[2,3] Zhixiang Fang,[2,3] and Xiaohan Yu[1]

[1]*Wuhan University of Technology, School of Computer Science and Technology, 122 Luoshi Road, Wuhan 430070, China*
[2]*Wuhan University, State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing,*
 129 Luoyu Road, Wuhan 430079, China
[3]*Wuhan University, Engineering Research Center for Spatio-Temporal Data Smart Acquisition and Application,*
 Ministry of Education of China, 129 Luoyu Road, Wuhan 430079, China

Correspondence should be addressed to Qingzhou Mao; qzhmao@whu.edu.cn

Saliency can be described as the ability of an item to be detected from its background in any particular scene, and it aims to estimate the probable location of the salient objects. Due to the salient map that computed by local contrast features can extract and highlight the edge parts including painting lines of Flying Apsaras, in this paper, we proposed an improved approach based on a frequency-tuned method for visual saliency detection of Flying Apsaras in the Dunhuang Grotto Murals, China. This improved saliency detection approach comprises three important steps: (1) image color and gray channel decomposition; (2) gray feature value computation and color channel convolution; (3) visual saliency definition based on normalization of previous visual saliency and spatial attention function. Unlike existing approaches that rely on many complex image features, this proposed approach only used local contrast and spatial attention information to simulate human's visual attention stimuli. This improved approach resulted in a much more efficient salient map in the aspect of computing performance. Furthermore, experimental results on the dataset of Flying Apsaras in the Dunhuang Grotto Murals showed that the proposed visual saliency detection approach is very effective when compared with five other state-of-the-art approaches.

1. Introduction

In Chinese traditional culture, Buddhism (particularly in the west of China) has played a prominent role in Buddhism history. This link is conveyed via the Dunhuang Grotto Murals, which have been of significant value to cultural studies across various Chinese dynasties. Protecting and maintaining the cultural integrity of the Dunhuang Grotto Murals have been a tremendous challenge for researchers. Digitalization, comprehensive preservation, and virtual interaction have been adopted as the key approaches to protecting and propagating the culture embodied in the Dunhuang Grotto Murals.

Flying Apsaras are a significant symbol of the Dunhuang art, and contain a lot of cultural information and connotations, which have been composed by a wide variety of people, namely, streamers, clouds totem, and so on. In addition,

because of the age, diverse colors, and complex composition of the images in Flying Apsaras, extracting the outlines accurately is of great significance in analyzing the painting styles of the Dunhuang Grotto Murals in different Chinese dynasties. Furthermore, due to the multioverlap phenomenon existing in some images of the Dunhuang Grotto Murals, that is, later works were painted directly on the early works, analyzing the cultural meanings of these images is even more difficult. Fortunately, one approach, called saliency detection, can be used to analyze the different painting styles of Flying Apsaras with different historical periods in the Dunhuang Grotto Murals. The relationship between Flying Apsaras and visual saliency can be closely connected by two important aspects: on the one hand, visual saliency can represent the initial state of these images without coloring, and saliency detection of Flying Apsaras in the Dunhuang Grotto Murals is of great historical value for historians in helping determine

the original state of the images in the Dunhuang Grotto Murals; on the other hand, the final salient map of Flying Apsaras that computed by local contrast features can extract and highlight the edge parts with striking local contrast; these parts including the painting lines have played a critical role in analyzing significant painting styles of Flying Apsaras. In addition, visual saliency detection can be used to extract the salient regions of Flying Apsaras, and these regions can represent the remarkable paintings' main thought, design, and plot.

One of the most challenging problems with perception is information overload. This overload presents a challenge when trying to obtain useful information from images due to the semantic gap between low level features and high level semantic concepts [1]. Visual attention is one of the primary features of human visual system (HVS) to derive important and compact information from the natural scenes [2]. And its mechanism enables a reduction of the redundant data that benefits perception during the selective attention process. In the image processing field, saliency detection is the process of detecting the interesting visual information. Furthermore, the selected stimuli need to be prioritized, with the most relevant being processed first and the less important ones later, and it leads to a sequential treatment of different parts of the visual scenes. Thus, saliency detection has been widely used in discovering regions of interests, predicting human visual attention, image quality assessment, and many other applications [3]. The saliency detection approach is used in many fields including object detection, object recognition, image enhancement, image rendering, image compression, and video summarization [4].

Many different approaches to visual saliency detection have been proposed by a variety of researchers from different research backgrounds. As pioneers, Koch, Ullman, and Poggio [5, 6] proposed a very influential Bioinspired model and Winner-Take-All selected mechanism to simulate visual system of humans. Itti et al. [7] derived bottom-up and top-down visual saliency using center-surround differences via multiscale features integration. Harel et al. [8] proposed a new bottom-up visual saliency detection model, graph-based visual saliency (GBVS), to form activation maps on certain feature channels and normalize them with highlighting conspicuity and combining with other maps. Liu et al. [9] used the conditional random field (CRF) approach to effectively combine a set of novel features including multiscale contrast, center-surround histogram, and color spatial distribution to describe salient object locally, regionally, and globally. Goferman et al. [10] proposed a new type of saliency model, context-aware saliency, which is based on four principles observed in the psychological literature: local low-level features, global considerations, visual organization rules, and high-level factors. Zhai and Shah [11] developed a fast approach for computing pixel-level salient maps using color histograms of images and a dynamic fusion technique applied to combine the temporal and spatial salient maps. Hou and Zhang [12] proposed a spectral residual approach for visual saliency detection to construct the corresponding salient maps in the spatial domain and extract the spectral residual of an image by analyzing the log-spectrum of it.

This is replaced by the phase spectrum of the Fourier transform in Gopalakrishnan et al. [13], because it is more effective and computationally efficient. Achanta et al. [14] combined low-level features of color and luminance to detect salient regions of images, which was called a frequency-tuned method. More recently, Yan et al. [15] employed a multilayer approach to analyze saliency cues and the final salient map was produced in a hierarchical model. Siva et al. [16] used an unsupervised learning approach that annotated objects of interest in images to capture the salient maps.

Due to employing only a few low-level features of images, the visual saliency detection results of the frequency-tuned method proposed by Achanta et al. [14] are not very satisfactory. In this paper, an alternative approach, histogram combined with image average and Gaussian blur (HIG), is proposed as an improvement to the frequency-tuned method. HIG combines features of color and luminance with the gray histogram information and uses the Euclidean distances between local features and global average features to define visual saliency as local contrast at pixel level. And the final visual saliency definition is based on normalization of prior visual saliency and spatial attention function. Unlike existing approaches that focus on multiple image features, the proposed approach only focuses on contrast and spatial attention information from the simulation of HVS, and these local contrast features have played a key role in highlighting the edge parts including painting lines of Flying Apsaras. This allows for more efficient computation of salient maps and very effective performance when compared to five other state-of-the-art approaches. These five state-of-the-art approaches of saliency detection are Zhai and Shah [11], Hou and Zhang [12], Achanta et al. [14], and Cheng et al. [17], hereby referred to as LC, SR, IG, HC, and RC, respectively. The choice of these algorithms was based on their frequency of citation in literature, popularity, and variety.

The rest of this paper is organized as follows: Section 2 provides an overview of the proposed saliency detection approach, followed by a more detailed description; Section 3 presents the experimental validations and comparison of results; finally, Section 4 contains the conclusion and recommendations for future works.

2. The Proposed Saliency Detection Approach

An overview of the improved saliency detection approach proposed for Flying Apsaras in the Dunhuang Grotto Murals in this paper is shown in Figure 1. This improved approach comprises the following three important steps: (1) image color and gray channel decomposition; (2) gray feature value and color channel convolution; (3) visual saliency definition based on normalization of prior visual saliency and spatial attention function. Using these three steps, the final salient map is determined from the input image that is shown as an example in Figure 1.

2.1. Decomposition of Color and Gray Components from the Input Images. It has been widely recognized among researchers that HVS does not equally process all
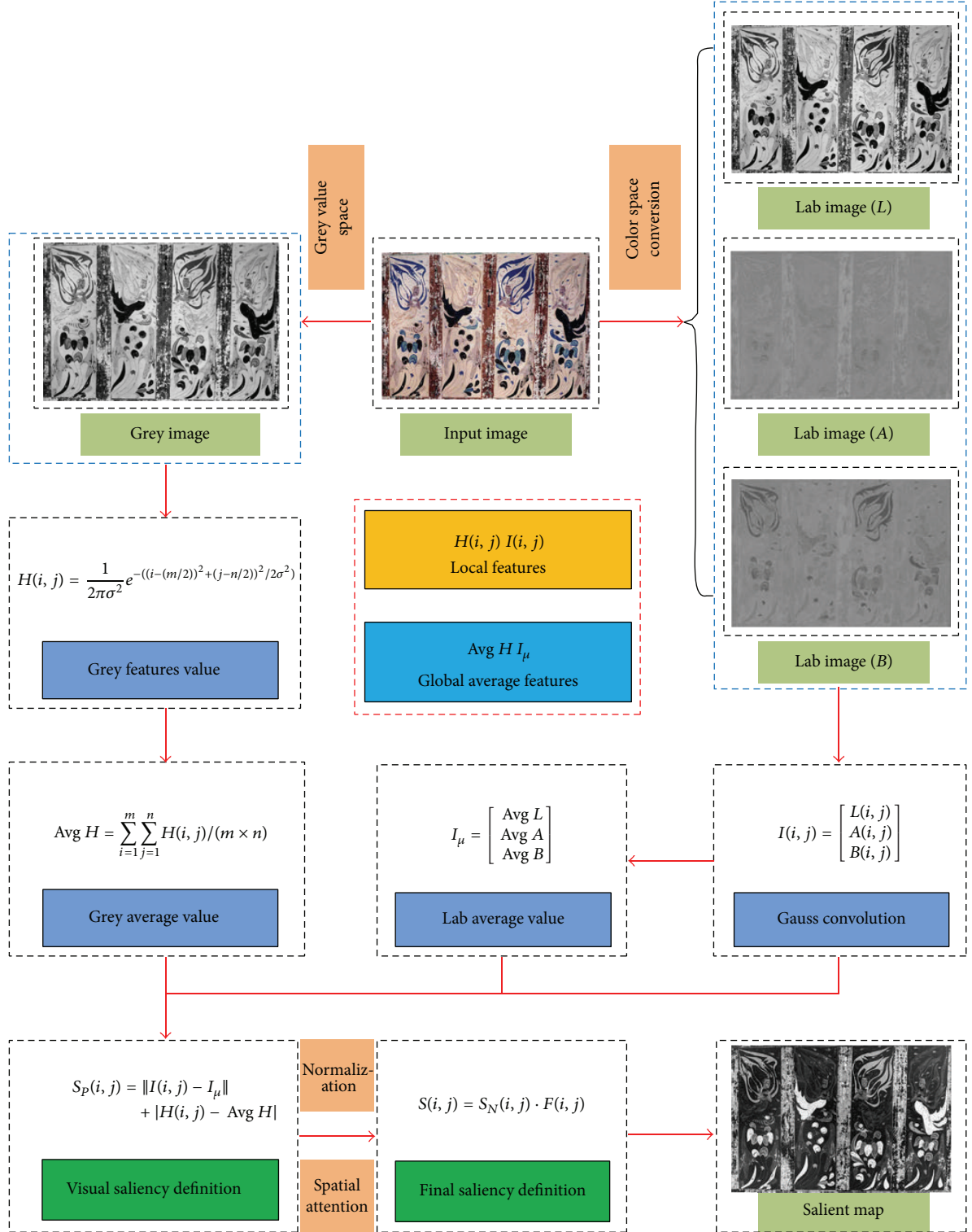
FIGURE 1: An overview of our saliency detection approach for Flying Apsaras in the Dunhuang Grotto Murals.

the information available to an observer. Color contrast has a significant impact on the ability to detect salient regions in images and has been highlighted in many previous works. For example, Osberger and Rohaly [18] suggest that a strong influence occurs when the color of a region is distinct from its background and some particular colors (e.g., red, yellow, and blue) attract our attention more than others.

For an RGB image, the three color channels are correlated linearly. In addition, because the value of each color channel is not related to the representation of stimulus intensity,

it is difficult to compute the difference in color intensity. Using a variety of color transformation formulas, other color spaces (e.g., HSI, Lab, and LUV, which are specified by the International Commission on Illumination (CIE)) can be derived from the RGB color space. The CIE's Lab color space is closest to human visual perception and perceptual uniformity, where L represents luminance information and matches human perception of lightness closely, a and b represent chromatic values, and the transformation formula can be shown in the following [19]:

$$X = 0.490 \times R + 0.310 \times G + 0.200 \times B,$$

$$Y = 0.177 \times R + 0.812 \times G + 0.011 \times B,$$

$$Z = 0.010 \times G + 0.990 \times B,$$

$$L = 116 f(Y) - 16, \tag{1}$$

$$a = 500 \cdot \left[f\left(\frac{X}{0.982} \right) - f(Y) \right],$$

$$b = 200 \cdot \left[f(Y) - f\left(\frac{Z}{1.183} \right) \right],$$

where

$$f(X) = \begin{cases} 7.787X + 0.138, & X \leq 0.008856 \\ X^{1/3}, & X > 0.008856. \end{cases} \tag{2}$$

In this paper, we utilize this nonlinear relational mapping of color channels L, a, and b and gray histogram information to imitate the nonlinear response of the human eyes. In addition, CIE's Lab is the most widely used and efficient image features for visual saliency detection algorithms, that is, researches in Itti et al. [7], Achanta et al. [20], Achanta et al. [14], Goferman et al. [10], Cheng et al. [17], and Yang et al. [21].

2.2. The Frequency-Tuned Saliency Detection Approach. Visual saliency refers to the ability of a vision system (human or machine) to select from an image a certain subset of visual information for further processing. This mechanism serves as a filter to select only the relevant information in the context of the current behaviors or tasks to be processed while ignoring other irrelevant information [4]. Achanta et al. [14] define the five key functions for a saliency detector as being able to (1) emphasize the largest salient objects; (2) uniformly highlight whole salient regions; (3) establish well-defined boundaries of salient objects; (4) disregard high frequencies arising from texture, noise, and blocking artifacts; and (5) efficiently output full resolution salient maps. These functions can be used when comparing alternative visual saliency detection approaches.

From the perspective of image frequency, IG approach [14] suggests that the input image can be divided into the constituent low frequency and high frequency parts of the frequency domain. The overall parts of structure information, such as contours and basic compositions of the original image, are reflected in the low frequency parts, while details

such as texture and noise are reflected in high frequency parts. In saliency detection parts, the overall structure information contained in low frequency parts has been widely used to compute visual saliency.

Let ω_{lc} and ω_{hc} represent the minimum frequency and the maximum frequency in the frequency domain, respectively. On the one hand, in order to accurately highlight the overall information representing potential salient objects, ω_{lc} should be low and it is helpful to highlight the entire object consistently and uniformly. On the other hand, in order to maintain the rich semantic information contained in any potential salient objects, ω_{hc} should be high, while remaining cognizant of the fact that the highest elements in the frequency domain should be discarded, for they can most likely be attributed to texture or noise parts.

Achanta et al. [14] chose the Difference of Gaussian (DoG) filter (3) for band-pass filtering to capture frequency limits ω_{lc} and ω_{hc}. This DoG filter has also been used for interest point detection [22] and saliency detection [7, 8]. The DoG filter is defined as

$$\begin{aligned} \text{DoG}(x, y) &= G(x, y, \sigma_1) - G(x, y, \sigma_2) \\ &= \frac{1}{2\pi} \left[\frac{1}{\sigma_1^2} e^{-(x^2+y^2)/2\sigma_1^2} - \frac{1}{\sigma_2^2} e^{-(x^2+y^2)/2\sigma_2^2} \right], \end{aligned} \tag{3}$$

where σ_1 and σ_2 ($\sigma_1 < \sigma_2$) are the standard deviations of the Gaussian function.

A DoG filter is a simple band-pass filter and its pass-band width is controlled by the ratio $\sigma_1 : \sigma_2$. The DoG filter is widely used in edge detection since it closely and efficiently approximates the Laplacian of Gaussian (LoG) filter, cited as the most satisfactory function for detecting intensity changes when the standard deviations of the Gaussian function are in the ratio $\sigma_1 : \sigma_2 = 1 : 1.6$ [23]. If we define $\sigma_1 = \rho\sigma$ and $\sigma_2 = \sigma$, namely, $\rho = \sigma_1/\sigma_2$, then the bandwidth of the DoG can be determined by ρ. Because a single DoG operation cannot cover the bandwidth $[\omega_{lc}, \omega_{hc}]$ several narrow band-pass DoG filters are combined and it is found that a summation over DoG with standard deviations can cover the bandwidth by the ratio ρ; the formula is given by

$$\begin{aligned} \sum_{n=0}^{N-1} \left[G\left(x, y, \rho^{n+1}\sigma \right) - G\left(x, y, \rho^n\sigma \right) \right] \\ = G\left(x, y, \rho^N\sigma \right) - G(x, y, \sigma), \end{aligned} \tag{4}$$

for an integer $N \geq 0$, which is the difference of two Gaussians whose standard deviations can have any ratio $K = \rho^N$. That is, we can obtain the combined result of applying several band-pass filters by choosing a DoG with a relatively large K. In order to make the bandwidth $[\omega_{lc}, \omega_{hc}]$ as large as possible, K must be large. In the extreme case $N = \infty$, then $G(x, y, \rho^N\sigma)$ is actually the average vector of the entire image. This process illustrates how the salient regions will be fully covered and not just focused on edges or in the center of the regions [14].

The main steps of IG approach are as follows: (1) the input RGB image I is transformed to a CIE's Lab color space; (2) the salient map S for the original image I is computed by

$$S(x, y) = \left\| I_{\omega_{hc}}(x, y) - I_\mu \right\|, \tag{5}$$

where I_μ is the arithmetic mean image feature vector, $I_{\omega_{hc}}(x, y)$ is a Gaussian blurred version of the original image using a 5×5 separable binomial kernel, $\|\cdot\|$ is the L_2 norm, and x and y are the pixel coordinates.

In conclusion, IG approach is a very simple visual saliency method to implement and only requires use of a Gaussian blur and image averaging process. From the point of view of implementation, IG approach is based on a global contrast method, and the global information is the mean feature vector of the original image.

2.3. The Improved Saliency Detection Approach. The saliency of an input image can be described as the probability that an item, distinct from its background, will be detected and output to a salient map in the form of an intensity map. The saliency detection algorithm aims to estimate the probable location of the salient object. A larger intensity value at a pixel indicates a higher saliency value, which means the pixel most probably belongs to the salient object. In Achanta et al. [14], features of color and luminance were exploited to define visual saliency. Specifically the L_2 norm between the corresponding image pixel vector value in the Gaussian blurred version and the mean image feature vector of the original image. In many natural scenes, particularly in the three widely used datasets including Bruce's [24] eye tracking dataset (BET), Judd's [25] eye tracking dataset (JET), and Achanta's [14] human-annotated dataset (AHA), the frequency-tuned saliency detection algorithm performs excellently. But, like many other saliency detection approaches, it is not effective for images of Flying Apsaras in the Dunhuang Grotto Murals. This is probably due to the rich content of the Dunhuang Grotto Murals, which presents a big challenge to distinguishing foregrounds from the mural's backgrounds.

In order to maintain the complete structural information of salient objects, as well as considering features of color and luminance, this paper also takes the gray histogram information of the original image and human visual attention stimuli into account and proposes an improved saliency detection approach for Flying Apsaras in the Dunhuang Grotto Murals. The gray feature contains rich luminance information and can be used to detect the salient objects of an image. In this paper, after converting the original image into gray-scale image, we can define the gray feature value $H(i, j)$ of pixel $p(i, j)$ as

$$H(i, j) = \frac{1}{2\pi\sigma^2} e^{-((i-m/2)^2 + (j-n/2)^2)/2\sigma^2}, \tag{6}$$

where σ is the normalized variance of Gaussian function and m and n represent the height and width of the input image I,

respectively. Thus, the mean value $\mathrm{Avg}\, H$ of gray feature can be expressed by

$$\mathrm{Avg}\, H = \frac{\sum_{i=1}^m \sum_{j=1}^n H(i, j)}{m \times n}, \tag{7}$$

where $m \times n$ represents the total number of pixels of the original image.

Meanwhile, after converting the input image in RGB color space into Lab color space, we can obtain the luminance value $L(i, j)$ and color feature values $A(i, j)$ and $B(i, j)$ of a pixel at the position (i, j). Then, using (3) to calculate the Gaussian burred version $I_{\omega_{hc}}(i, j) = (GYL(i, j), GYA(i, j), GYB(i, j))^T$ of the original image, as expressed in Achanta's research, the salient regions will be fully covered and not just be highlighted on their edges or in the center of the salient regions. Then, the arithmetic mean image feature vector $I_\mu = (\mathrm{Avg}\, L, \mathrm{Avg}\, A, \mathrm{Avg}\, B)^T$ of the input image can be determined by

$$\mathrm{Avg}\, L = \frac{\sum_{i=1}^m \sum_{j=1}^n L(i, j)}{m \times n}, \qquad \mathrm{Avg}\, A = \frac{\sum_{i=1}^m \sum_{j=1}^n A(i, j)}{m \times n},$$

$$\mathrm{Avg}\, B = \frac{\sum_{i=1}^m \sum_{j=1}^n B(i, j)}{m \times n}. \tag{8}$$

Combining features of color and luminance with the gray histogram information, this paper exploits the visual contrast principle to define visual saliency as local contrast based on pixel level. We consider that pixels with significant difference compared with the mean vector are part of the salient regions and assume that pixels with values close to the mean vector are part of the background. Therefore, the saliency $S_P(i, j)$ of a pixel at location (i, j) is characterized by the difference between it and the mean vector; namely,

$$S_P(i, j) = \left\| I_{\omega_{hc}}(i, j) - I_\mu \right\| + \left| H(i, j) - \mathrm{Avg}\, H \right|, \tag{9}$$

where $\|\cdot\|$ is the L_2 norm and $|\cdot|$ is the absolute value function.

Because the final salient map is actually a gray image, in order to optimize display of the final salient map, it can be processed by a gray-scale transformation based on some normalized functions. In this paper, the normalized salient map $S_N(i, j)$ can be further computed by

$$S_N(i, j) = 255 \times \frac{S_P(i, j) - S_{\min}}{S_{\max} - S_{\min}}, \tag{10}$$

where $S_{\max}$ and $S_{\min}$ denote the maximum and minimum values of the salient map $S_P(i, j)$, respectively. Normalization ensures these values of $S_N(i, j)$ have a unified range from 0 to 255 for different image locations, which is convenient for visualization of final output.

Generally when an observer views an image, he or she has a tendency to stare at the central location without making eye movements. This phenomenon is referred to as central effect [26]. In Zhang et al. [26], when searching the scenes for a conspicuous target, fixation distributions are shifted from

the image center to the distributions of image features. Based on the center shift mechanism, in this paper, we define a spatial attention function $F(i, j)$, a weight decay function, to reflect the spatial attention tendency of human eyes as follows:

$$F(i, j) = \frac{L - 2D(i, j)}{L + 2D(i, j)}, \qquad (11)$$

where $D(i, j)$ denotes the distance between pixel at location (i, j) and the center of a given image and L is the length of the diagonal of the input image. For any pixel at location (i, j), $D(i, j) \in [0, L/2]$; thus, $F(i, j) \in [0, 1]$, and $F(i, j)$ is a monotonically decreasing function with the variable $D(i, j)$.

In summary, the definition of the spatial attention function is reasonable, since it does not exclude the possibility for the pixels located at the boundary of an image to be noticed [26]. In order to integrate the extracted features and interaction effects of human observations that are measured by the spatial attention function $F(i, j)$, the final salient map in the proposed approach is evaluated by

$$S(i, j) = F(i, j) \cdot S_N(i, j). \qquad (12)$$

This improved saliency detection approach of combining gray histogram information, features of color and luminance, and the human's spatial visual attention function for solving visual saliency detection problem of the Flying Apsaras in the Dunhuang Grotto Murals can be expressed by Algorithm 1.

Algorithm 1. The improved saliency detection approach for Flying Apsaras in the Dunhuang Grotto Murals, China, is as follows.

> *Input*: Images of Flying Apsaras in the Dunhuang Grotto Murals, China.
>
> *Output*: Visual saliency detection maps of images in the Dunhuang Grotto Murals, China.
>
> *Process*:
>
> (1) Convert the original image into a grayscale image, and then compute the gray feature value $H(i, j)$ of pixel $p(i, j)$ according to the gray feature function according to (6). Obtain the mean value Avg H of gray feature simultaneously.
>
> (2) Convert the original RGB color space image into CEI's Lab color space format to obtain three different property values $L(i, j)$, $A(i, j)$, and $B(i, j)$ of three different color channels L, a, and b.
>
> (3) Use a 5×5 separable binomial kernel of pixel at the position to obtain the Gaussian blurred version $I_{\omega_{hc}}$ of the original image, and obtain the arithmetic mean image feature vector I_μ of input image simultaneously.
>
> (4) Define the new saliency detection formula shown in (9), and use (10) to export the normalized salient map $S_N(i, j)$.
>
> (5) Integrate the spatial attention function $F(i, j)$ to the normalized salient map $S_N(i, j)$ to obtain the final definition of visual saliency detection $S(i, j)$.
>
> (6) Export the final visual saliency detection maps of the original images.

3. Experimental Results

In order to evaluate the performance of this proposed visual saliency detection approach, an image dataset of Flying Apsaras in the Dunhuang Grotto Murals was established from Zheng and Tai's [27] book. To prove the effectiveness of the proposed approach, this paper compares experimental results obtained using the proposed approach with those obtained when applying five other state-of-the-art algorithms.

3.1. The Dataset of Flying Apsaras in the Dunhuang Grotto Murals. This paper collected the experimental data (300 images) of the Flying Apsaras in the Dunhuang Grotto Murals published in Zheng and Tai's [27] book. These pictures cover four historical periods: the beginning stage (Northern Liang (421–439 A.D.), Northern Wei (439–535 A.D.), and Western Wei (535–556 A.D.)), the developing stage (Northern Zhou (557–581 A.D.), Sui Dynasty (581–618 A.D.), and Early Tang (618–712 A.D.)), the flourishing stage (High Tang (712–848 A.D.), Late Tang (848–907 A.D.), and Five Dynasties (907–960 A.D.)), and the terminal stage (Song Dynasty (960–1036 A.D.), Western Xia (1036–1227 A.D.), and Yuan Dynasty (1227–1368 A.D.)).

3.2. Saliency Detection Results of Four Historical Periods. All the algorithms are used to compute the salient maps for the same dataset. Our approach is implemented in MATLAB R2009a software environment. The other algorithms are downloaded from the authors' homepages. All these algorithms were implemented using Windows XP with an Intel(R) Core (TM) i3-2130 CPU@ 3.40 GHz processor and 4 GB of memory. Figure 2 shows that the salient maps produced by the proposed algorithm can highlight cultural elements such as streamers, clouds totem, and clothes of Flying Apsaras; particularly for images in the flourishing stage, the line structure and spatial layout that reflect painting styles are highlighted for further processing, that is, image enhancement or image rendering. This is beneficial to the objective of preserving the complete structural information of salient objects and it is useful for researchers when analyzing the spatial structure of streamers with different periods.

3.3. Comparison with Five Other State-of-the-Art Approaches. Visual saliency detection aims to produce the salient maps of images via simulating the behavior of HVS. The proposed approaches compared with five state-of-the-art approaches designed for saliency detection, including HC [17], LC [11], RC [17], SR [12], and IG [14]. Visual saliency ground truth is constructed from eye fixation patterns of 35 observers in our experiments. The state-of-the-art performance of visual saliency detection for real scene images with fixation ground truth similar to the case was reported by Judd et al. [25].

Figure 3 illustrates that the salient area can be adequately distinguished from the background area via the proposed approach. One reason for this is that the contrast based on color component and gray histogram information is
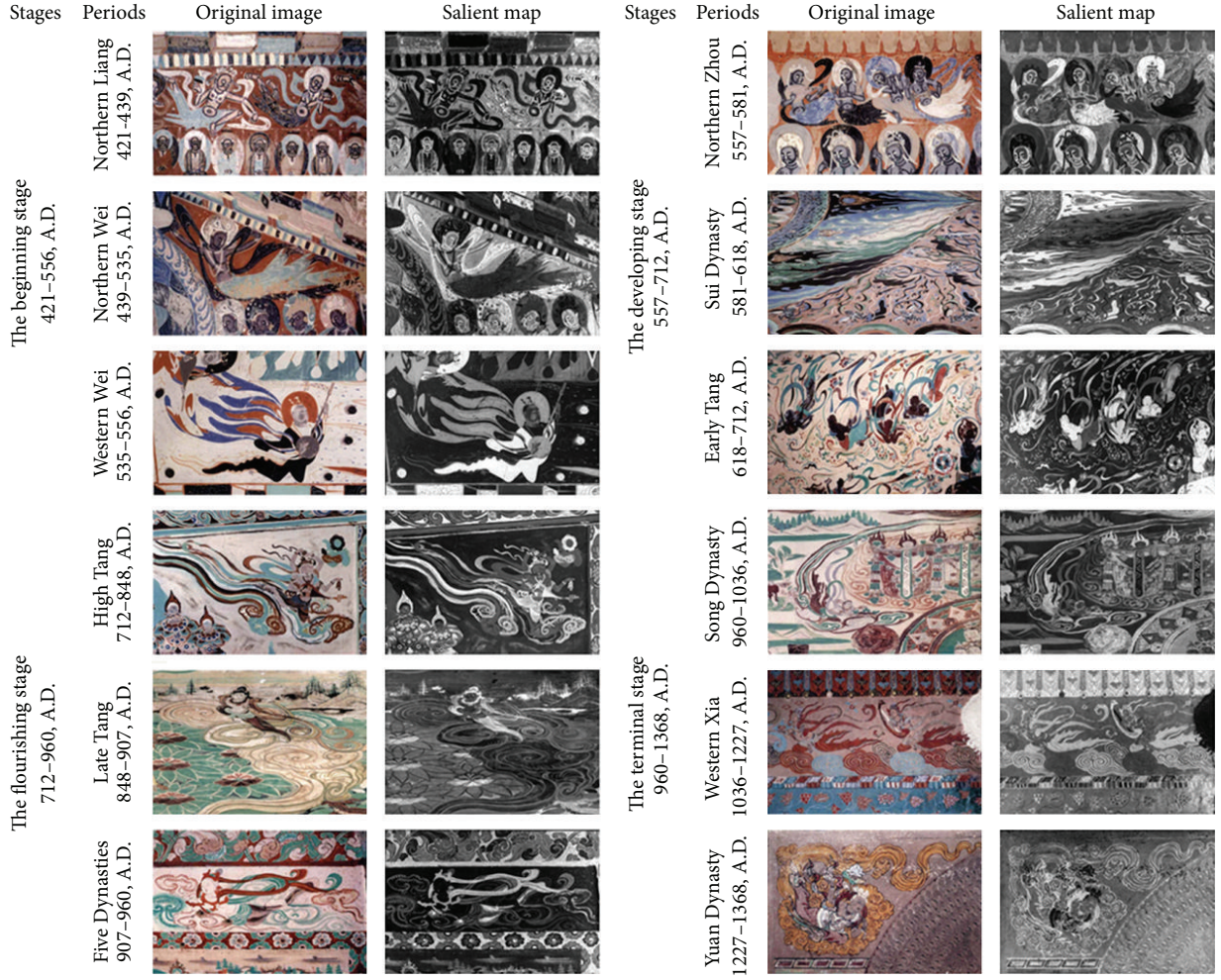
Figure 2: Visual saliency detection results of Flying Apsaras in the Dunhuang Grotto Murals within four historical periods: the beginning stage, the developing stage, the flourishing stage, and the terminal stage.

more distinguishable than other representations; the other reason is that a spatial attention function to simulate human's visual attention stimuli is adopted in our approach. From perspective of highlighting painting lines of Flying Apsaras, the final salient maps by using of the proposed approach are more consistent than the other approaches, with the ability to highlight the painting lines of Flying Apsaras' streamers in the Dunhuang Grotto Murals and guarantee their integrity simultaneously. And from Figure 3, it can be concluded that the saliency detection results obtained by using the proposed approach (HIG) are outperforming results obtained using the frequency-tuned method (IG).

3.4. Performance Evaluation of Different Saliency Detection Approaches.

In order to evaluate the pros and cons of each approach in a quantitative way, a False Positive Rate and True Positive Rate (FPR-TPR) curve, which is a standard technique for information for our Flying Apsaras datasets retrieval community, is adopted in this paper. False Positive Rate and True Positive Rate (or Recall) are defined as [28–30]

$$\text{False Positive Rate} = \frac{\text{FP}}{\text{FP} + \text{TN}},$$

$$\text{True Positive Rate} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{13}$$

where FP is false positive (regarding the pixels not in the obtained salient regions (background) as the pixels in the ground truth), TN is true negative (regarding the pixels not in the obtained salient regions (background) as the pixels not in the ground truth), TP is true positive (regarding the pixels in the obtained salient regions (foreground) as the pixels in the ground truth), and FN is false negative (regarding the pixels in the obtained salient regions (foreground) as the pixels not in the ground truth). False Positive Rate is the probability that a true negative was labeled a false positive, and True Positive Rate corresponds to the fraction of the detected salient pixels belonging to the salient object in the ground truth.

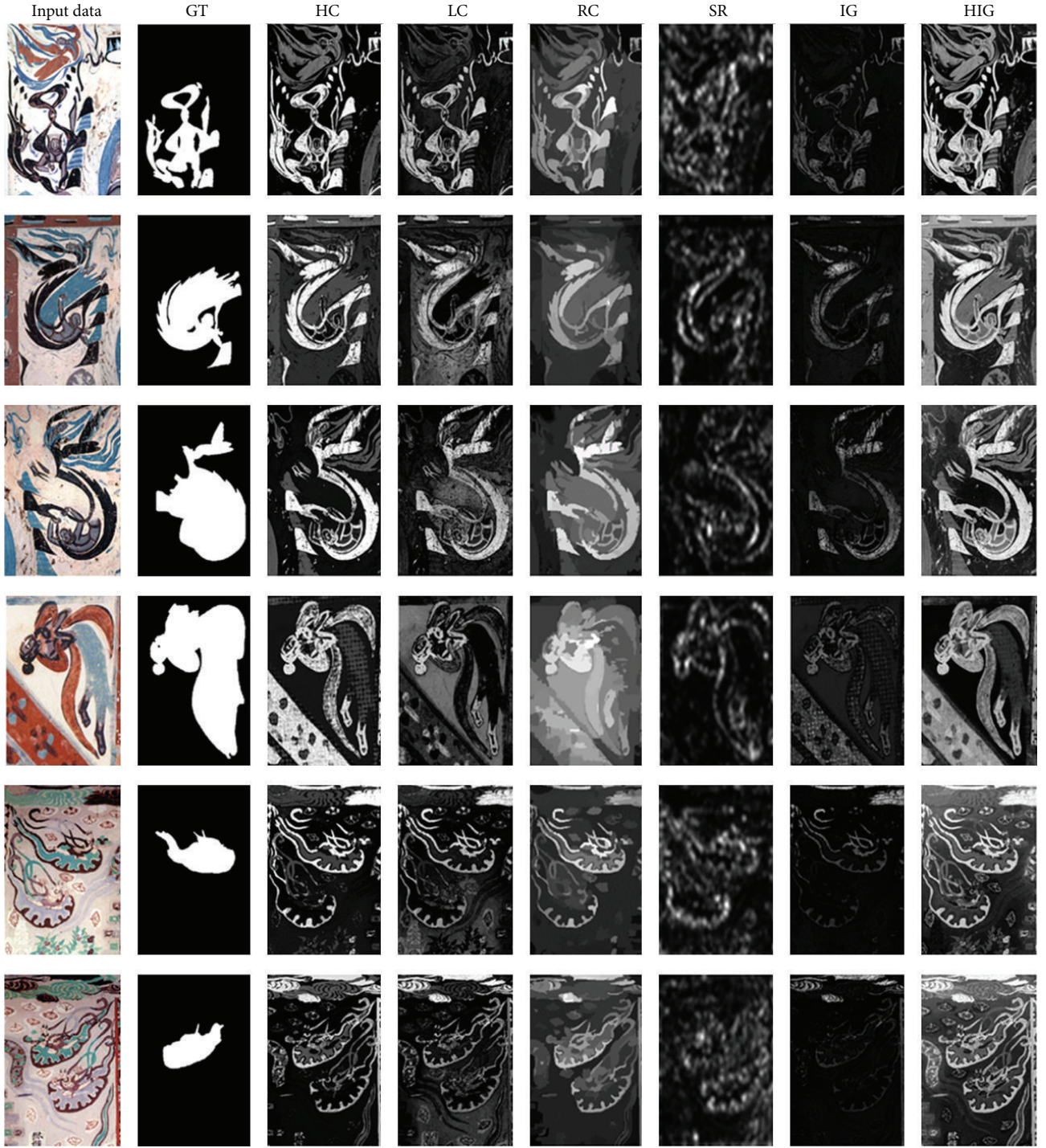| Input data	GT	HC	LC	RC	SR	IG	HIG

FIGURE 3: Comparison of visual detection results via different algorithms: HC [17], LC [11], RC [17], SR [12], IG [14], and HIG. And all these algorithms are compared based on the visual saliency ground truth.

Using MABLAB software, the FPR-TPR curves were plotted for the different saliency detection algorithms and shown in Figure 4. Our experiment follows the settings in [14, 17], where the salient maps are digitized at each possible threshold within the range [0, 255]. Our approach achieves the highest TPR in almost the entire FPR range [0, 1]. This is because combining saliency information from three scales

makes the background generally have low saliency values. Only sufficiently salient objects can be detected in this case.

From these FPR-TPR curves in Figure 4, we can see that our approach has better performances when there is less FPR; thus, it can be concluded that the improved saliency detection approach (HIG) is superior to the other five state-of-the-art approaches. Achieving the same TPR, the proposed
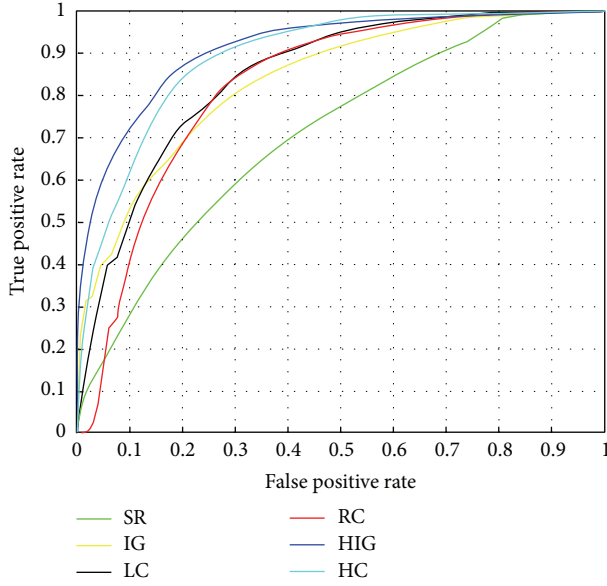
FIGURE 4: FPR-TPR curves for different visual saliency detection algorithms: HC [17], LC [11], RC [17], SR [12], IG [14], and HIG in the experiments.

FIGURE 5: Statistical comparson results (Precision, Recall, F-measure, and AUC) of different visual saliency detection algorithms: HC [17], LC [11], RC [17], SR [12], IG [14], and HIG in the experiments.

algorithm has the lowest FPR, which means it can detect more salient regions; and with the same FPR, the proposed algorithm has high TPR, which means it can detect salient regions more accurately.

In addition, Precision, Recall (or True Positive Rate), F-measure, and the values of area under the FPR-TPR curves (AUC) are also used to evaluate performances of these saliency detection approaches quantitatively. Recall is defined as above, and Precision is defined as [14]

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{14}$$

Precision corresponds to the percentage of detected salient pixels correctly assigned, while Recall corresponds to the fraction of the detected salient pixels belonging to the salient object in the ground truth. High Recall can be achieved at the expense of reducing Precision. So, it is necessary and important to measure them together; therefore, F-measure is defined as [14]

$$F\text{-measure} = \frac{\left(1 + \beta^2\right) \times \text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}}. \tag{15}$$

We use $\beta^2 = 0.3$ suggested in Achanta et al. [14] to weight Precision more than Recall.

In Figure 5, compared with the state-of-the-art results, we can concluded that the rank of these approaches is HIG > HC [17] > LC [11] > IG [14] > RC [17] > SR [12], and our improved approach (HIG) shows the highest Precision (90.13%), Recall (94.59%), F-measure (91.12%), and AUC (91.51%) values in statistical sense, which are a little higher than HC [17] with the Precision (84.52%), Recall (93.82%), F-measure (86.50%), and AUC (89.39%) values, and this conclusion also can be supported by Figure 4.
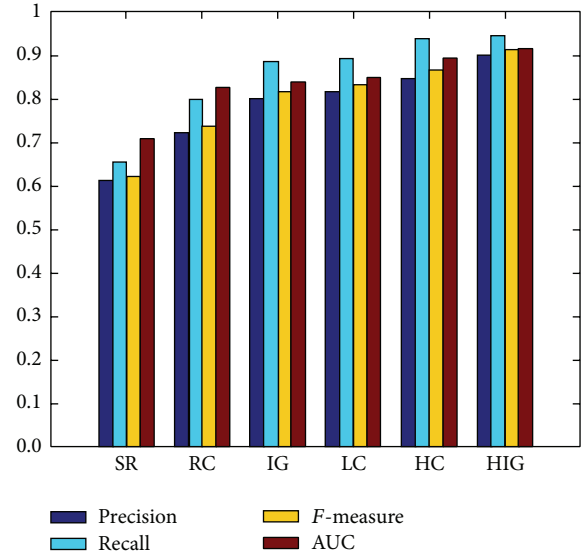
4. Conclusion and Future Works

In this paper, an improved saliency detection approach based on a frequency-tuned method for Flying Apsaras in the Dunhuang Grotto Murals has been proposed. This proposed approach utilizes a CIE's Lab color space information and gray histogram information. In addition, a spatial attention function to simulate human's visual attention stimuli is proposed, which is easy to implement, and provides consistent salient maps. Furthermore, it is beneficial to highlight the painting lines of Flying Apsaras' streamers in the Dunhuang Grotto Murals and guarantee their integrity simultaneously. The final experiment results on the dataset of Flying Apsaras in the Dunhuang Grotto Murals show that the proposed approach is very effective when compared to the five state-of-the-art approaches. In addition, the quantitative comparison results demonstrated that the proposed approach outperforms the five classical approaches in terms of FPR-TPR curves, Precision, Recall, F-measure, and AUC.

Future research will consider the use of deep learning techniques to extract semantic information from the Dunhuang Murals, and based on the extracted semantic information, visual saliency detection approaches should be enhanced. More recently, these approaches including Deep Belief Networks (DBN), Restricted Boltzmann Machine (RBM), and Convolutional Neural Networks (CNN) could be applied to determine various activities in the Dunhuang Grotto Murals, for example, the depicted musical entertainment scene. Recognition of these scenes would be highly valuable in ongoing efforts to analyze Chinese cultural backgrounds and evolution in the early dynasties between 421 A.D. and 1368 A.D.

Conflict of Interests

Acknowledgments

References

[1] F. Xu, L. Zhang, Y. Zhang, and W. Ma, "Salient feature selection for visual concept learning," in *Advances in Multimedia Information Processing—CM 2005*, vol. 3767 of *Lecture Notes in Computer Science*, pp. 617–628, Springer, Berlin, Germany, 2005.

[2] N. Imamoğlu, W. Lin, and Y. Fang, "A saliency detection model using low-level features based on wavelet transform," *IEEE Transactions on Multimedia*, vol. 15, no. 1, pp. 96–105, 2013.

[3] Z. Chen, J. Yuan, and Y. P. Tan, "Hybrid saliency detection for images," *IEEE Signal Processing Letters*, vol. 20, no. 1, pp. 95–98, 2013.

[4] A. Borji, D. N. Sihite, and L. Itti, "Salient object detection: a benchmark," in *Proceedings of the European Conference on Computer Vision*, pp. 414–429, 2012.

[5] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," in *Matters of Intelligence*, pp. 115–141, Springer, Amsterdam, The Netherlands, 1987.

[6] C. Koch and T. Poggio, "Predicting the visual world: silence is golden," *Nature Neuroscience*, vol. 2, no. 1, pp. 9–10, 1999.

[7] 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.

[8] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Advances in Neural Information Processing Systems*, vol. 19, pp. 545–552, 2007.

[9] T. Liu, Z. Yuan, J. Sun et al., "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353–367, 2011.

[10] 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.

[11] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proceedings of the 14th Annual ACM International Conference on Multimedia (MULTIMEDIA '06)*, pp. 815–824, October 2006.

[12] 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.

[13] 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, Miami, Fla, USA, June 2009.

[14] R. Achanta, S. Hemamiz, F. Estraday, and S. Süsstrunky, "Frequency-tuned salient region detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '09)*, pp. 1597–1604, June 2009.

[15] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 1155–1162, IEEE, Portland, Ore, USA, June 2013.

[16] P. Siva, C. Russell, T. Xiang, and L. Agapito, "Looking beyond the image: unsupervised learning for object saliency and detection," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 3238–3245, Portland, Ore, USA, June 2013.

[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] W. M. Osberger and A. M. Rohaly, "Automatic detection of regions of interest in complex video sequences," in *6th Human Vision and Electronic Imaging*, vol. 4299 of *Proceedings of SPIE*, pp. 361–372, San Jose, Calif, USA, June 2001.

[19] C. Connolly and T. Fleiss, "A study of efficiency and accuracy in the transformation from RGB to CIE Lab color space," *IEEE Transactions on Image Processing*, vol. 6, no. 7, pp. 1046–1048, 1997.

[20] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk, "Salient region detection and segmentation," in *Proceedings of the International Conference on Computer Vision Systems*, pp. 66–75, 2008.

[21] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*, pp. 3166–3173, June 2013.

[22] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[23] D. Marr, *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information*, W. H. Freeman, San Francisco, Calif, USA, 1982.

[24] N. D. B. Bruce and J. K. Tsotsos, "Saliency based on information maximization," in *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS '05)*, pp. 155–162, December 2005.

[25] 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, IEEE, Kyoto, Japan, October 2009.

[26] H. Zhang, W. Wang, G. Su, and L. Duan, "A simple and effective saliency detection approach," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR '12)*, pp. 186–189, November 2012.

[27] R. Zheng and J. Tai, *Collected Edition of Flying Apsaras in the Dunhuang Grotto Murals*, Commercial Press, Hong Kong, 2002, (Chinese).

[28] K. Bowyer, C. Kranenburg, and S. Dougherty, "Edge detector evaluation using empirical ROC curves," *Computer Vision and Image Understanding*, vol. 84, no. 1, pp. 77–103, 2001.

[29] I. E. Abdou and W. K. Pratt, "Quantitative design and evaluation of enhancement/thresholding edge detectors," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 753–763, 1979.

[30] D. R. Martin, C. C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 530–549, 2004.