Signal Processing: Image Communication 69 (2018) 1–7

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

A novel superpixel-based saliency detection model for 360-degree images

Yuming Fang [a,*], Xiaoqiang Zhang [a], Nevrez Imamoglu [b]

[a] School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330032, Jiangxi, China
[b] National Institute of Advanced Industrial Science and Technology, Tokyo, Japan

ARTICLE INFO

Keywords:
Visual attention
360-degree image
Saliency detection
Figure-ground law
Boundary connectivity
Gestalt theory

ABSTRACT

Effective visual attention modeling is a key factor that helps enhance the overall Quality of Experience (QoE) of VR/AR data. Although a huge number of algorithms have been developed in recent years to detect salient regions in flat-2D images, the research on 360-degree image saliency is limited. In this study, we propose a superpixel-level saliency detection model for 360-degree images by figure-ground law of Gestalt theory. First, the input image is segmented into superpixels. CIE Lab color space is then used to extract the perceptual features. We extract luminance and texture features for 360-degree images from L channel, while color features are extracted from a and b channels. We compute two components for saliency prediction by figure-ground law of Gestalt theory: feature contrast and boundary connectivity. The feature contrast is computed on superpixel level by luminance and color features. The boundary connectivity is predicted for background measure and it describes the spatial layout of image region with two image boundaries (upper and lower boundary). The final saliency map of 360-degree image is calculated by fusing feature contrast map and boundary connectivity map. Experimental results on a public eye tracking database of 360-degree images show promising performance of saliency prediction from the proposed method.

1. Introduction

The Human Visual System (HVS) can rapidly and accurately process visual information in its visual field [1]. With large amount of visual information, selective attention in the HVS would select the most relevant parts while ignoring others when viewing visual scenes. Various studies in computer vision have tried to build computational models of visual attention by stimulating the attention mechanisms in the HVS [2]. Saliency detection models can be generally categorized as bottom-up and top-down approaches. Bottom-up mechanism is a data-driven, and task-independent perception process for automatic salient region selection for natural scenes [3–14], while top-down mechanism is a task-dependent processing affected by the performed tasks, feature distribution of targets, etc. [15–19].

During the past decades, many saliency detection models have been successfully used in various visual processing applications such as retargeting [20], quality assessment [16,21], coding [22,23], segmentation [24], *etc*. Itti et al. proposed a computational model of visual attention based on the neuronal architecture of the primates' early visual system [14]. In that study, multi-scale feature contrasts from intensity, color and orientation are extracted to compute the saliency map [14]. Based on Itti's model, a graph-based saliency detection model was proposed by using a new measure of dissimilarity [12]. Brue et al. adopted the concept of information maximization for visual attention modeling [25]. That visual attention model measures Shannon's self-information to obtain the saliency map. The characteristics of the HVS by including contrast visual masking, perceptual decomposition, sensitivity functions, and center–surround interactions are used for visual attention modeling [26]. Later, Liu et al. proposed a novel saliency detection model by designing a saliency tree [27].

Another type of saliency detection models are built in transform domain. The concept of spectral residual in transform domain was proposed for saliency detection [6]. In that study, Hou et al. used the log spectra representation of images in Fourier Transform to calculate saliency value for images. Later, phase spectrum was used to design a visual attention model for video sequences in [22]. Guo et al. predicted saliency map by Inverse Fourier Transform on the original phase spectrum and the constant amplitude spectrum [22]. Chen et al. presented a frequency-based saliency model based on the Fourier Transformation of multiple spatial Gabor filters [28].

Based on learned sparse codes, Wang et al. presented a saliency detection model by site entropy rate [11]. Recently, some researchers adopted patch-level contrast for visual attention modeling [20,29]. Goferman et al. designed a saliency detection model by considering the

* Corresponding author.
E-mail addresses: fa0001ng@e.ntu.edu.sg (Y. Fang), yzzxiaoqiang@qq.com (X. Zhang), nevrez.imamoglu@aist.go.jp (N. Imamoglu).

https://doi.org/10.1016/j.image.2018.07.009
Received 11 September 2017; Received in revised form 25 July 2018; Accepted 25 July 2018
Available online 1 August 2018

Fig. 1. 360-degree image sample. Left boundary and right boundary of 360-degree image are connected. The rightmost sun and the leftmost sun belong to the same sun.

context of the dominant objects as well as the objects themselves [29]. Zhang et al. presented an eye fixation prediction model by surroundedness prediction [30]. Fang et al. introduced saliency detection models in compressed domain for images and video sequences [20,31]. Qi et al. proposed Graph-Boolean Map method to predict salient regions [32]. In that study, Qi et al. use a graph inference based on belief propagation to generate saliency map.

These saliency detection methods mentioned above are all designed for flat-2D images or video. Recently, virtual reality (VR) has been growing increasingly popular in both industry and academia, where 360-degree images are widely used in the related VR applications. Therefore, it is meaningful to design effective visual attention models for 360-degree images, which can be used in various VR applications such as the Quality of Experience of VR. Different from flat-2D images, 360-degree images only have two boundaries (upper and lower boundary) as shown in Fig. 1, while left boundary and right boundary of 360-degree images are connected.

Recently, ICME 2017 has held a competition named Salient360! Grand Challenge to try to solve the problem of 360-degree images saliency detection [33]. Participants were provided with a series of 360-degree images with given ground truth data. The details on the dataset are described by Rai et al. [34]. Some models presented in this competition try to deal with the 360-degree saliency detection [35–41]. In [35], Ling et al. extracted the image features by dictionary based on sparse representation. Then image features and latitude-bias enhancement were used to estimate 360-degree image saliency. In [36], the authors explored how existing visual saliency models for 2D images can be extended to saliency detection models of 360-degree images in the equirectangular format. Different from the methods mentioned above which only use low-level features, Battisti et al. combined low-level and semantic features for visual saliency estimation [39].

Currently, there are a few saliency detection models exploiting the boundary prior [31,32,42–44]. However, all of these existing studies use the boundary prior for flat-2D image saliency detection. Some of these models directly assume the image boundary as background [43,44]. In [42], only regions that connect to image boundary are likely to be background regions, while others will be considered as foreground regions. Recently, some superpixel-level saliency detection models such as [45,46] have been proposed to try to improve the saliency detection performance on flat-2D images.

In this study, based on figure-ground law of Gestalt theory, we propose a novel saliency detection model for 360-degree image. The figure-ground law of Gestalt theory states that foreground is more smaller and has high contrast compared to the background [47]. According to this statement, we calculate feature contrast and boundary connectivity for image saliency computation. First, the 360-degree image is segmented into superpixels. The color space is then converted to CIE Lab color space, in which we extract the luminance and texture features from channel L, while color features are extracted from a and b channels. Superpixel-level feature contrast is computed on each channel in Lab space and boundary connectivity is estimated for background measure. Finally, we combine the boundary connectivity and feature contrast together to predict the final saliency map of 360-degree image.

The contributions of this paper: (1) Based on figure-ground law of Gestalt theory, we propose a novel saliency detection model for 360-degree images. (2) A reliable background measure, called 360-degree boundary connectivity, is defined to compute background prior in 360-degree images.

The remaining of this paper is organized as follows. In Section 2, the proposed model is described in detail. Section 3 provides the experimental results on the eye tracking database. The final section concludes the paper.

2. Proposed model

In this study, we propose a novel saliency detection model based on Gestalt theory for 360-degree image. The proposed framework is shown in Fig. 2. It firstly segments 360-degree image into superpixels. Then we convert the color space into CIE Lab color space. The luminance, texture, and color features are extracted from each segmented channel images. Based on the figure-ground law of Gestalt theory, the feature contrast is calculated by different features and boundary connectivity is defined to measure background. The final saliency map is computed by combining the boundary connectivity and feature contrast. We will introduce the details step by step in the following subsections.

2.1. Superpixel segmentation

Some studies use superpixel segmentation to segment images as a preprocessing step in visual attention modeling [45,46]. Compared with pixel-level methods, superpixel-level methods can reduce redundant partial image information, and provide more useful spatial structure for saliency detection.

Here, we use the simple linear iterative clustering (SLIC) algorithm [48] to segment 360-degree images into superpixels. SLIC algorithm segments an input image using K-means clustering in RGBXY space, which can yield local and edge-aware superpixels. By default, the number k of clustering centers is the only parameter of the algorithm. First, k initial cluster centers $M_i = (L_i, a_i, b_i, x_i, y_i)$ are sampled on a regular grid spaced $G = \sqrt{N/k}$ where N is the pixel number of input image. After the initialization, K-means clustering operation is performed to estimate superpixels.

2.2. Feature extraction

First, we convert the color space into CIE Lab color space, in which luminance and texture features are extracted from the channel L, and

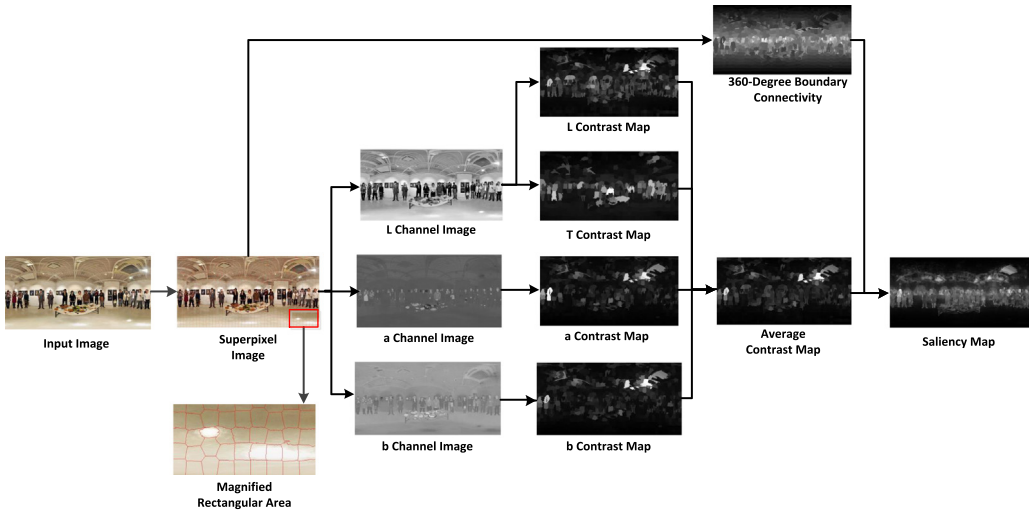

Fig. 2. The proposed framework.

color features are extracted from a and b channels. The figure-ground law of Gestalt theory claims that foreground is smaller and with high contrast compared to background [47]. With these features, the feature contrast C_i^n of superpixel i is calculated based on Gestalt theory as:

$$C_i^n = -\sum_{i \neq j} D_{ij}^n log(N(d_{ij})) \tag{1}$$

where $n \in \{L, T, a, b\}$; D_{ij} denotes the feature difference between image superpixels i and j and it is calculated by $L1$ norm; N is the normalization operation; d_{ij} is geodesic distance between superpixels i and j. By using a weighting factor based on the distance d_{ij}, we put more emphasis for the feature contrast computation on those regions closer to the current patch. The geodesic distance d_{ij} of 360-degree image superpixels is computed as:

$$d_{ij} = \begin{cases} \sqrt{x^2 + y^2} & x \leq w/2 \\ \sqrt{(w-x)^2 + y^2} & x > w/2 \end{cases} \tag{2}$$

where x and y are horizontal distance and vertical distance between superpixels i and j, respectively; w is width of 360-degree image. Texture feature T can be computed as follows:

$$T = \sqrt{T_a^2 + T_b^2} \tag{3}$$

where T_a and T_b are horizontal gradient map and vertical gradient map based on L channel.

Finally, four feature contrast maps C^n are calculated in three channels including L, a, and b channel by Eq. (1). The overall feature contrast map $\bar{C}$ of the input 360-degree image is computed as the average of the feature contrast:

$$\bar{C} = \frac{1}{M} \sum_{n \in \{L,T,a,b\}} N(C^n) \tag{4}$$

where N is the normalization operator and M is the number of features ($M = 4$).

2.3. 360-degree boundary connectivity

In this study, we also use image boundary information for saliency calculation based on Gestalt theory. The figure-ground law of Gestalt theory indicates that foreground is more smaller and with high contrast compared to background [47]. Here, a novel and reliable background measure, called 360-degree boundary connectivity is defined to obtain background prior in 360-degree image by figure-ground law of Gestalt theory. The definition is shown as follows.

$$F_i = \frac{N(\alpha_i)}{1 + \lambda N(\beta_i)} \tag{5}$$

where F_i is 360-degree boundary connectivity of superpixel i; α_i is minimum distance between superpixel i and 360-degree image boundaries (upper and lower boundary); β_i represents size of superpixel i; λ is a parameter. As shown in Eq. (5), the boundary connectivity is proportional to the distance and inversely proportional to the superpixel size. The figure-ground law of Gestalt theory suggests that small regions generally belong to foreground [47]. A region that is more smaller and farther away from the boundaries, the more likely it belongs to foreground.

2.4. Saliency prediction

As described above, we calculate the feature contrast map and boundary connectivity map for any given 360-degree image. The saliency of the image can be predicted based on these two maps by:

$$S = \bar{C} * F \tag{6}$$

where $\bar{C}$ is average feature contrast map; F is 360-degree boundary connectivity map. By combining these two maps, we can adaptively obtain image saliency for each input 360-degree image.

Furthermore, we incorporate multiple scales to improve the contrast between salient and non-salient regions. Specifically, the saliency map $\bar{S}$ for an input image can be calculated by:

$$\bar{S} = \frac{1}{R} \sum_{k \in \{r, 2r, 3r, 4r\}} S^k \tag{7}$$

where r denotes the number of superpixels for the input image; R is the number of scales ($R = 4$); $k \in \{r, 2r, 3r, 4r\}$; S^k is defined in Eq. (6).

We show some saliency map samples at different scales in Fig. 3. From this figure, we can see that the saliency map from scale r detects the details of the background, while the saliency map at multi-scale is more smooth than those at single scale.

Harel et al. observed that most of fixation nodes concentrate on center of image [12]. There are two reasons for this: first, photographers often use salient objects as the center of image; second, it is natural for an observer to focus directly on the center of image when viewing an image [12]. For 360-degree images [49], although observer can view 360-degree images freely (the body moves back and forth, or the head moves up and down), there is also a similar center bias for these two reasons as shown below. Therefore, we have added the center bias as post-processing.

Inspired by [39,50], obtaining the clustering of the fixation points by convolving the fixation points map with a Gaussian kernel, is usually used to produce the saliency map. The saliency map S convolving with a Gaussian kernel as follows:

$$S' = G_\sigma * S \tag{8}$$

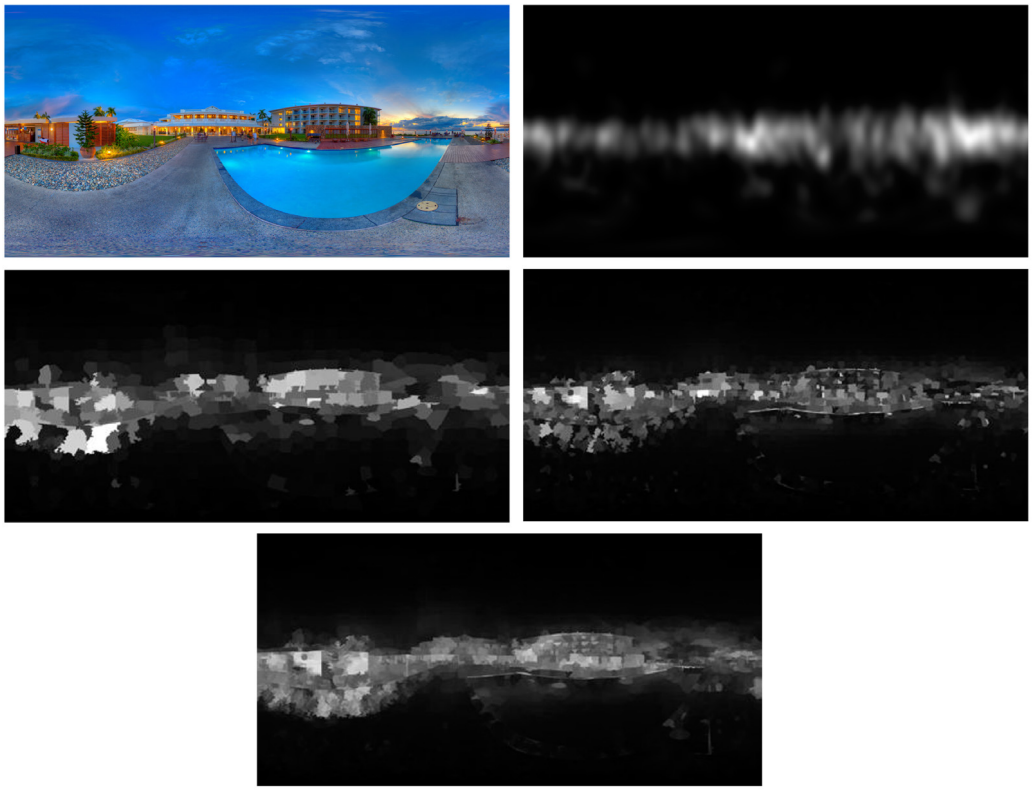

Fig. 3. Multi-scale saliency map samples. Row 1: original image and ground-truth, respectively; row 2: saliency maps from scale r and scale $4r$, respectively; row 3: saliency maps from multi-scale.

where G_σ is Gaussian kernel function with σ as the standard deviation of the Gaussian filter; S' is final saliency map.

3. Experimental results

In this section, we conduct the comparison experiments to evaluate the performance of the proposed method. We first introduce the evaluation methodology and quantitative evaluation metrics. Then the performance of the proposed method is evaluated by comparing with other existing ones.

3.1. Evaluation methodology

We perform comparison experiments on Salient360 database [34,49] including 45 360-degree images with eye tracking data. Salient360 database contains many indoor or outdoor 360-degree images with complex backgrounds and a variety of small objects. Generally, an effective saliency detection model would predict the saliency maps similar with the fixation maps. Three commonly used evaluation metrics are used to evaluate the performance of the proposed method: the Receiver Operating Characteristics (ROC), the Linear Correlation Coefficient (CC) and the Kullback–Leibler divergence (KL) [51].

ROC curve [51] is widely used for performance evaluation of saliency detection models in the literature. The saliency map from computational saliency detection model can be divided into salient points and non-salient points through defining a threshold. The fixation map from eye tracking data includes the target points and the background points. The True Positive Rate (TPR) is defined as the percentage of target points falling into the salient points from a computational saliency detection model, while the False Positive Rate (FPR) is defined as the percentage of background points falling into the salient points. The ROC curve of a specified saliency detection model can be obtained as the curve of TPR vs. FPR through choosing different thresholds. Specifically, TPR and FPR are defined as:

$$TPR = \frac{TP}{TP + FN} \tag{9}$$

$$FPR = \frac{FP}{FP + TN}. \tag{10}$$

Generally, the higher the ROC curve is, the better the predication performance of the saliency detection model is for the saliency prediction. The area under ROC curve (AUC) can be used to measure the saliency performance quantitatively. With larger AUC values, the performance is better for saliency prediction.

CC [51] provides a measure of degree of linear correlation between the saliency map and fixation map. In this experiment, we use Pearson's correlation coefficient between two variables to calculate CC for the saliency detection models. It is defined as the covariance of the saliency map and fixation map divided by the product of their standard deviations as following:

$$CC(s, f) = \frac{cov(s, f)}{\sigma_s \sigma_f} \tag{11}$$

where s and f denote saliency map and fixation map, respectively; $cov(s, f)$ is the covariance; σ_s and σ_f are the standard deviation values of the saliency map and fixation map, respectively. The correlation coefficient is computed in the range of [0, 1], where 0 implies that there is no linear correlation between the saliency map and fixation map, while 1 implies a linear equation of the relationship between the saliency map and fixation map. With larger CC value, the better the performance of the saliency detection model is.

Besides ROC and CC, KL [51] is also used to evaluate the performance of the proposed method in this study. KL is used to measure the distance between these two distributions as following.

$$KL(A, B) = \frac{1}{2}\left(\sum_n a_n log \frac{a_n}{b_n} + \sum_n b_n \frac{b_n}{a_n}\right) \tag{12}$$

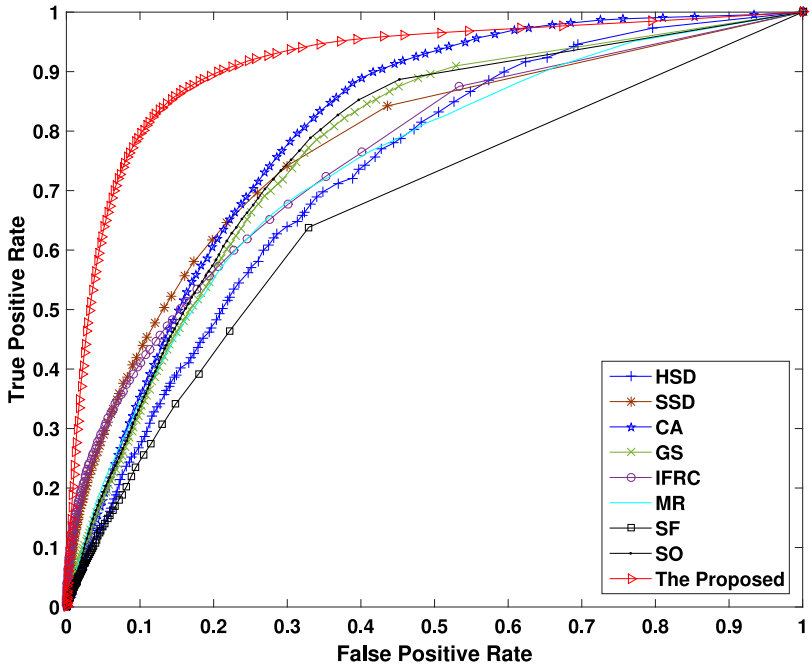

Fig. 4. The comparison results of ROC from different models.

Table 1
AUC and CC values from different saliency detection models.

Models	HSD	SSD	CA	GS	IFRC	MR	SF	SO	The proposed
AUC	0.72	0.77	0.82	0.77	0.75	0.74	0.59	0.76	0.91
CC	0.38	0.42	0.47	0.40	0.43	0.39	0.18	0.38	0.73

where A and B are saliency distribution at human fixation locations and random locations with probability density functions a_n and b_n, respectively. and n is the index of the saliency value bin. With larger KL value, the better the performance of the saliency detection model is.

3.2. Comparison with other existing related works

In this study, we use the following saliency detection models to conduct the comparison experiment: HSD [52], SSD [53], CA [29], GS [54], IFRC [50], MR [44], SF [8], and SO [42]. HSD [52] presents saliency detection model by a measure of region scales, which is compatible with human perception of object scales and helps deal with small salient objects. SSD [53] proposes a visual saliency method in the frequency domain, where the authors assume that the secret of visual saliency may mainly hide in the phases of intermediate frequencies. In CA [29], the salient regions are regarded to contain not only dominant objects but also the parts of the background. GS [54] exploits a new background measure to provide more clues for saliency detection. IFRC [50] estimates visual saliency by nonlinearly integrating features using region covariance. In MR [44], saliency detection is considered as a ranking and retrieval problem. In SF [8], local and the global contrast are both considered in a unified way to estimate visual saliency. SO [42] investigates the boundary prior to assist other saliency clues for visual saliency estimation.

We provide the quantitative results in Table 1, where CC and AUC results from different saliency detection models are provided. Those scores are all averaged over images in the database Saliency360. From this table, we can see that the proposed model obtain much higher values of CC and AUC than those of the other studies, which demonstrates that the proposed model can predict much more accurate saliency results than other studies.

Furthermore, The ROC curves of these compared models are provided in Fig. 4, to evaluate the performance of the proposed saliency detection model. From this figure, we can see that the proposed model can provide the best performance among the compared models.

Additionally, we can observe the comparison results of different image saliency detection models in Fig. 5. From the second and fourth rows, we can see that HSD [52] and CA [29] detect some background regions as salient. For SSD [53] and IFRC [50], the main problem is that both of them directly assume the image boundaries as backgrounds. As shown in the third and sixth rows in Fig. 5, only center regions are detected as salient in images. From the fifth, seventh and ninth rows of Fig. 5, we can see that GS [54], MR [44] and SO [42] fail to detect salient regions accurately due to the complex backgrounds of images. For SF [8], it may lose some salient information in images as shown in the eighth row. Overall, compared with the ground truth in the last row of Fig. 5, the proposed model can predict more accurate saliency results than other existing ones, as shown in the saliency maps of the tenth row in Fig. 5.

3.3. Comparison with other 360-degree saliency models

The Table 2 shows the experimental results of the proposed model and other saliency detection models on the evaluation dataset of Salient360! Challenge. The existing models used in the comparison experiment contain WHU [35], SJTU [37], TU1 [38], TU3 [38], TU5 [38], URome3 [39], CDSR [35], ZJU [36] (denoted as 'BMS-FSM-Eq' in the original manuscript) and DCT. As shown in this table, TU Munich [38] submitted several different models and the performance of these models is listed in this table. Wuhan Univ. [35] submitted two models to the Salient360! Challenge and CDSR [35] is the model presented in that paper. DCT was submitted to the Salient360! Challenge by ourselves. From this table, we can see that the proposed model outperforms other

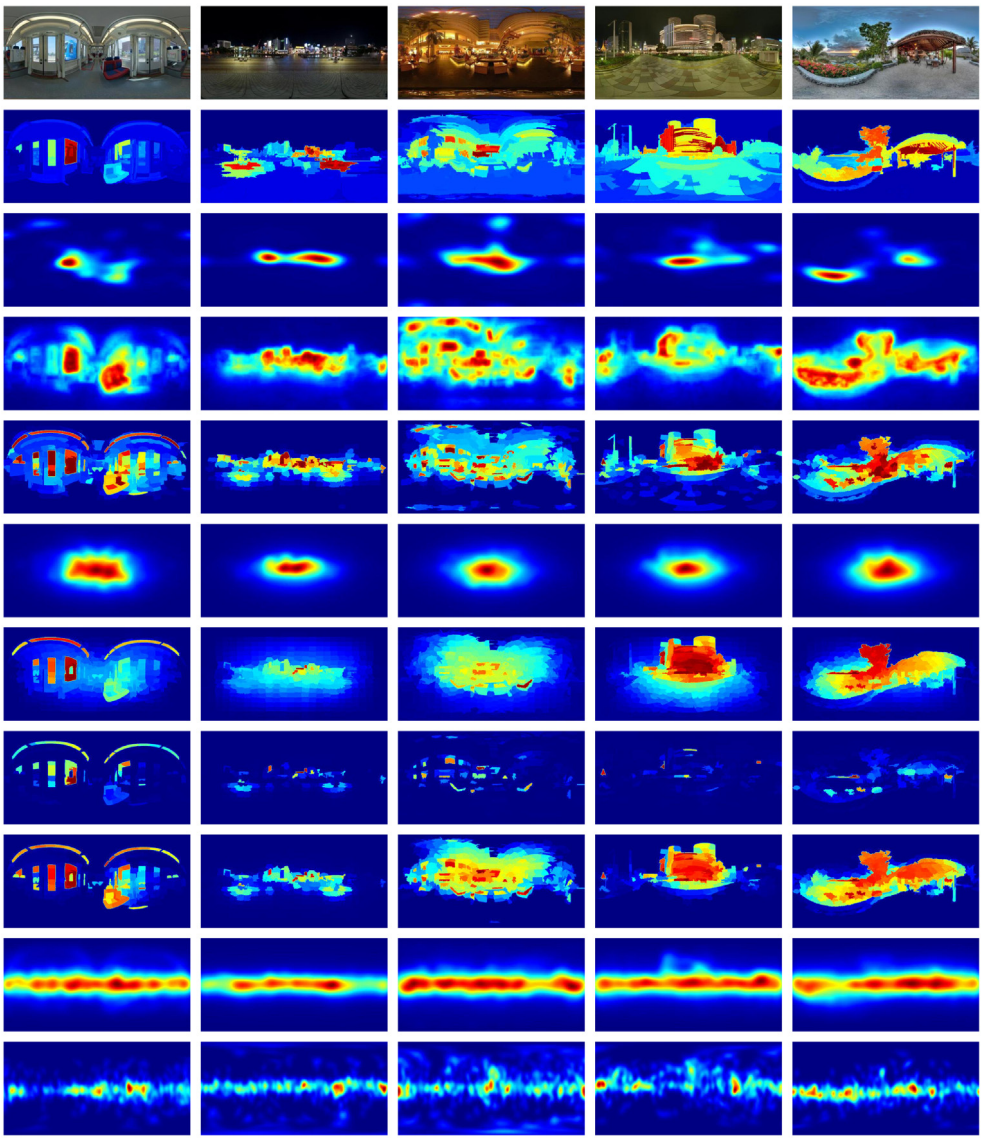

Fig. 5. Sample saliency maps. Row 1: original images; Row 2–11: saliency maps from HSD [52], SSD [53], CA [29], GS [54], IFRC [50], MR [44], SF [8], SO [42], and the proposed model, and the ground truth map.

Table 2
Comparison of the proposed model with other saliency prediction models presented in 360-degree images Salient360! Challenge (DCT denotes the method we submitted to this Challenge and it is different from the proposed model) on evaluation dataset.

Models	WHU	SJTU	TU1	TU3	TU5	URome3	CDSR	DCT	ZJU	The proposed
KL	0.51	0.65	0.75	0.74	0.64	0.81	0.50	1.14	0.44	0.80
CC	0.71	0.67	0.62	0.60	0.56	0.52	0.73	0.57	0.69	0.75

models in terms of CC values on head motion. For KL values, the proposed method can obtain better performance than URome3 and DCT models (salient360! submission version). The experimental results demonstrate that the proposed saliency detection model by Gestalt theory can obtain promising performance in saliency prediction of 360-degree images.

4. Conclusion

In this paper, we propose a novel saliency detection model based on Gestalt theory for 360-degree image. According to figure-ground law of Gestalt theory, the superpixel-level feature contrast in each channel in CIE Lab space is adopted and the concept of 360-degree boundary connectivity is defined as background measure to extract background map. The final saliency map for 360-degree image is predicted by combining the feature contrast and boundary connectivity. Experimental results show that the proposed saliency detection model can obtain promising performance for 360-degree image saliency prediction.

However, the proposed model is not the best one by KL values compared with other 360-degree saliency models presented in Salient360! Challenge. The reason may be that it ignores other important features, such as semantic features. In terms of efficiency, the computational complexity can be further improved. In the future, we will further explore the visual saliency detection for 360-degree images in the aspects of effectiveness and efficiency.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61571212, the Fok Ying-Tong Education Foundation, China under Grant 161061, and by the Natural Science Foundation of Jiangxi Province, China under Grant GJJ160420, 20171BBE50068, and 20171BCB23048.

References

[1] S.E. Palmer, Visual Science: Photons to Phenomenology, MIT Press, 1999.
[2] A. Borji, L. Itti, State-of-the-art in visual attention modeling, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 185–207.
[3] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2009, pp. 1597–1604.
[4] K.Y. Chang, T.L. Liu, H.T. Chen, S.H. Lai, Fusing generic objectness and visual saliency for salient object detection, in: International Conference on Computer Vision, 2011, pp. 914–921.
[5] V. Gopalakrishnan, Y. Hu, D. Rajan, Random walks on graphs for salient object detection in images, IEEE Trans. Image Process. 19 (12) (2010) 3232–3242.
[6] X. Hou, L. Zhang, Saliency detection: A spectral residual approach, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
[7] Y. Fang, Z. Wang, W. Lin, Z. Fang, Video saliency incorporating spatiotemporal cues and uncertainty weighting, IEEE Trans. Image Process. 23 (9) (2014) 3910–3921.
[8] F. Perazzi, P. Krahenbuhl, Y. Pritch, A. Hornung, Saliency filters: Contrast based filtering for salient region detection, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2012, pp. 733–740.
[9] L. Wang, J. Xue, N. Zheng, G. Hua, Automatic salient object extraction with contextual cue, in: International Conference on Computer Vision, 2011, pp. 105–112.
[10] Y. Fang, C. Zhang, J. Li, J. Lei, M.P.D. Silva, P.L. Callet, Visual attention modeling for stereoscopic video: A benchmark and computational model, IEEE Trans. Image Process. 26 (10) (2017) 4684–4696.
[11] W. Wang, Y. Wang, Q. Huang, W. Gao, Measuring visual saliency by site entropy rate, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2010, pp. 2368–2375.
[12] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, in: Advances in Neural Information Processing Systems, 2007, pp. 545–552.
[13] S. Li, C. Zeng, Y. Fu, S. Liu, Optimizing multi-graph learning based salient object detection, Signal Process., Image Commun. 55 (2017) 93–105.
[14] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysisk, IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (1998) 1254–1259.
[15] Y. Fang, W. Lin, C.T. Lau, B.-S. Lee, A visual attention model combining top-down and bottom-up mechanisms for salient object detection, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, 2011, pp. 1293–1296.
[16] Z. Lu, W. Lin, X. Yang, E. Ong, S. Yao, Modeling visual attention's modulatory aftereffects on visual sensitivity and quality evaluation, IEEE Trans. Image Process. 14 (11) (2005) 1928–1942.
[17] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, H. Shum, Learning to detect a salient object, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2) (2011) 353–367.
[18] J. Yang, M.H. Yang, Top-down visual saliency via joint crf and dictionary learning, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2296–2303.
[19] A. Torralba, A. Oliva, M.S. Castelhano, J.M. Henderson, Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search, Psychol. Rev. 113 (4) (2006) 766–786.
[20] Y. Fang, Z. Chen, W. Lin, C.-W. Lin, Saliency detection in the compressed domain for adaptive image retargeting, IEEE Trans. Image Process. 21 (9) (2012) 3888–3901.
[21] W. Lin, C.-C.J. Kuo, Perceptual visual quality metrics: A survey, J. Vis. Commun. Image Represent. 22 (4) (2011) 297–312.
[22] C. Guo, L. Zhang, A novel multi-resolution spatiotemporal saliency detection model and its applications in image and video compression, IEEE Trans. Image Process. 19 (1) (2010) 185–198.
[23] Z. Pan, J. Lei, Y. Zhang, X. Sun, S. Kwong, Fast motion estimation based on content property for low-complexity H.265/HEVC encoder, IEEE Trans. Broadcast. 62 (3) (2016) 675–685.
[24] J. Lei, H. Zhang, L. You, C. Hou, L. Wang, Evaluation and modeling of depth feature incorporated visual attention for salient object segmentation, Neurocomputing 120 (9) (2013) 24–33.
[25] N.D.B. Bruce, J.K. Tsotsos, Saliency based on information maximization, Adv. Neural Inf. Process. Syst. 18 (3) (2006) 155–162.
[26] O. Le Meur, P. Le Callet, D. Barba, D. Thoreau, A coherent computational approach to model the bottom-up visual attention, IEEE Trans. Pattern Anal. Mach. Intell. 28 (5) (2006) 802–817.
[27] Z. Liu, W. Zou, O. Le Meur, Saliency tree: A novel saliency detection framework, IEEE Trans. Image Process. 23 (5) (2014) 1937–1952.
[28] D. Chen, T. Jia, C. Wu, Visual saliency detection: From space to frequency, Signal Process., Image Commun. 44 (2016) 57–68.
[29] S. Goferman, L. Zelnik-Manor, A. Tal, Context-aware saliency detection, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2010, pp. 1915–1926.
[30] J. Zhang, S. Sclaroff, Exploiting surroundedness for saliency detection: A boolean map approach, IEEE Trans. Pattern Anal. Mach. Intell. 38 (5) (2016) 889–902.
[31] Y. Fang, W. Lin, Z. Chen, C.-M. Tsai, C.-W. Lin, A video saliency detection model in compressed domain, IEEE Trans. Circuits Syst. Video Technol. 24 (1) (2014) 27–38.
[32] W. Qi, J. Han, Y. Zhang, L.-F. Bai, Graph-boolean map for salient object detection, Signal Process., Image Commun. 49 (2016) 9–16.
[33] Salient360!: Visual attention modeling for 360° images grand challenge, 2017. http://www.icme2017.org/grand-challenges/.
[34] Y. Rai, J. Gutiérrez, P. Le Callet, A dataset of head and eye movements for 360 degree images, in: ACM on Multimedia Systems Conference, 2017, pp. 205–210.
[35] J. Ling, K. Zhang, Y. Zhang, D. Yang, Z. Chen, A saliency prediction model on 360 degree images using color dictionary based sparse representation, Signal Process., Image Commun. 69 (2018) 60–68.
[36] P. Lebreton, A. Raake, GBVS360, BMS360, ProSal: Extending existing saliency prediction models from 2D to omnidirectional images, Signal Process., Image Commun. 69 (2018) 69–78.
[37] Y. Zhu, G. Zhai, X. Min, The prediction of head and eye movement for 360 degree images, Signal Process., Image Commun. 69 (2018) 15–25.
[38] M. Startsev, M. Dorr, 360-aware saliency estimation with conventional image saliency predictors, Signal Process., Image Commun. 69 (2018) 43–52.
[39] F. Battisti, S. Baldoni, M. Brizzi, M. Carli, A feature-based approach for saliency estimation of omnidirectional images, Signal Process., Image Commun. 69 (2018) 53–59.
[40] M.A. Reina, K. McGuinness, X. Giro-i Nietro, N.E. O'Connor, Scanpath and saliency prediction on 360 degree images, Signal Process., Image Commun. 69 (2018) 8–14.
[41] R. Monroy, S. Lutz, T. Chalasani, A. Smolic, SalNet360:' Saliency maps for omni-directional images with CNNs, Signal Process., Image Commun. 69 (2018) 26–34.
[42] W. Zhu, S. Liang, Y. Wei, J. Sun, Saliency optimization from robust background detection, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2814–2821.
[43] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, S. Li, Salient object detection: A discriminative regional feature integration approach, in: IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 251–268.
[44] C. Yang, L. Zhang, H. Lu, X. Ruan, M.-H. Yang, Saliency detection via graph-based manifold ranking, in: IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3166–3173.
[45] Z. Liu, W. Zou, O.L. Meur, Saliency tree: A novel saliency detection framework, IEEE Trans. Image Process. 23 (5) (2014) 1937–1952.
[46] J. Lei, B. Wang, Y. Fang, W. Lin, P.L. Callet, N. Ling, C. Hou, A universal framework for salient object detection, IEEE Trans. Multimed. 18 (9) (2016) 1783–1795.
[47] M. Soegaard, The Laws of Figure/Ground, Prägnanz, Closure, and Common Fate — Gestalt Principles (3), UX design, 2016.
[48] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, SLIC superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell. 34 (11) (2012) 2274–2282.
[49] Y. Rai, P. Le Callet, P. Guillotel, Which saliency weighting for omni directional image quality assessment? in: Quality of Multimedia Experience, QoMEX, 2017, pp. 1–6.
[50] E. Erdem, A. Erdem, Visual saliency estimation by nonlinearly integrating features using region covariances, J. Vision 13 (4) (2013) 11–11.
[51] J. Gutiérrez, E. David, Y. Rai, P. Le Callet, Toolbox and Dataset for the Development of Saliency and Scanpath Models for Omnidirectional/360° Still Images, Elsevier, 2018.
[52] J. Shi, Q. Yan, L. Xu, J. Jia, Hierarchical image saliency detection on extended CSSD, IEEE Trans. Pattern Anal. Mach. Intell. 38 (4) (2016) 717–729.
[53] J. Li, L.-Y. Duan, X. Chen, T. Huang, Y. Tian, Finding the secret of image saliency in the frequency domain, IEEE Trans. Pattern Anal. Mach. Intell. 37 (12) (2015) 2428–2440.
[54] Y. Wei, F. Wen, W. Zhu, J. Sun, Geodesic saliency using background priors, in: European Conference on Computer Vision, 2012, pp. 29–42.