Visual Saliency Detection Via Background Features and Object-Location Cues

Muwei Jian[1, 2]
[1]*School of Computer Science and Technology, Shandong University of Finance and Economics,China.*
[2]*School of Creative Technologies, University of Portsmouth, Portsmouth, UK*
jianmuwei@ouc.edu.cn

Jing Wang
Department of Computer Science and Technology, Ocean University of China, Qingdao, China.
1731005693@qq.com

Hui Yu
School of Creative Technologies, University of Portsmouth, Portsmouth, UK
hui.yu@port.ac.uk

Yakun Ju
Department of Computer Science and Technology, Ocean University of China, Qingdao, China.
1104380083@qq.com

***Abstract*—In this paper, we propose a simple visual saliency-detection model based on spatial position of salient objects and background cues. At first, discrete wavelet frame transform (DWDT) are used to extract directionality characteristics for estimating the centoid of salient objects in the input image. Then, the colour contrast feature performed is to represent the physical characteristics of salient objects. Conversely, sparse dictionary learning is applied to obtain the background feature map. Finally, three typical cues of the directional feature, the colour contrast feature and the background feature are mixed to generate a credible saliency map. Experimental results verify that the designed method is useful and effective.**

Keywords—discrete wavelet transform, saliency detection, background features, position prior

I. Introduction

This Visual saliency detection is one of the hot research issues in computer vision, digital signal processing, pattern recognition and etc. Generally, visual saliency detection frameworks mainly fit into two categories: one is the bottom-up saliency-detection model, which is primarily engaged in employing and extracting of the underlying features in the input image/video, such as colour, texture, intensity, contrast, direction, motion [1, 2].

In 1998, Itti et al. [3] proposed an early saliency-detection model, which used the center surround algorithm to obtain the colour, intensity, and orientation salient maps of the original image. These salient feature maps were merged to obtain a final saliency map. In [4], Harel and Koch improved the Itti's saliency-detection method and proposed a bottom-up graph-based visual saliency-detection model (GBVS) by forecasting human eyes' fixations. Hou and Zhang [5] devised a saliency detection framework using frequency analysis of spectral residuals. In this model, saliency maps were obtained via fourier transform by calculating the characteristic spectrum of the input image. In [6], Hou et al. proposed a novel spatio-temporal saliency detection algorithm based on phase spectrum of quaternion fourier transform to target saliency regions in an image or video. A weakness of frequency analysis based models is that the boundary of the significant object or target is possibly not clear and the salient regions cannot be clearly highlighted. In [8], Achanta et al. proposed a saliency-detection method for content-aware image resizing, which utilized salient colour and intensity information to preserve salient objects in the original image. Rahtu et al. [9] proposed a saliency-detection approach to estimate of saliency values through conditional random field (CRF), which yielded satisfactory results on natural images and video. In [10], Liu et al. designed a saliency-detection model based on the conditional random field (CRF). In their model, three characteristics of colour space distribution, central-surrounding contrast and multi-scale contrast were combined to produce a saliency map. Goferman et al. [11] proposed a novel context-based saliency detection model based on four psychological observations to extract important regions that represent the scenes. In [12], an efficient saliency detection method was proposed to predict saliency values by using the clues of the middle and bottom layers within a Bayesian framework. Yang et al. [13] devised a saliency-detection system based on graph-based ranking, which utilized visual information of the upper, lower, left and right borders of the image, as a priori position distribution of the background and foreground, to obtain a saliency map. In [14], Cheng et al. proposed an effective saliency detection model based on global contrast (GC) and region contrast (RC). Liu et al. [15] devised a deep hierarchical saliency network based on convolutional neural networks for salient object detection.

In contrast, the top-down saliency detection model is application-oriented approach that is generally a supervised learning procedure requiring of huge amounts of labelled data. Thereby this type of saliency detection models is not particularly large. In [16], a top-down saliency detection method based on global scene configuration was proposed for object detection (e.g. a person). Kanan et al. [17] designed a top-down model based on the appearance of salient objects through a Bayesian framework, which can better predict human fixations. In [18], Cholakkal et al. proposed a top-down method by containing coupled image classification blocks and a class-aware sparse coding strategy for salient object detection.

The rest of the paper is organized as follows. In Section II, we will introduce the proposed framework in detail. Experimental results are presented in Section III. The paper closes with a conclusion and discussion in Section IV.

This work This work was supported by National Natural Science Foundation of China (NSFC) (61601427, 61602229, 61771230); Royal Society-K. C. Wong International Fellowship (NIF\R1\180909); Mount Tai Youth Talents Plan.

II. The Proposed Method

A. Spatial position center prior

In this section, we first present the proposed bottom-up saliency detection framework in detail. The spatial position of salient object can be seen as a reliable and forward visual feature for HVS during saliency detection [19]. To reckon with the spatial location of the salient object, discrete wavelet frame transform (DWDT) are applied to represent the multiple directional features of the salient object.

Compared with traditional discrete fourier transform (DFT), discrete wavelet transform (DWT) and discrete wavelet frame transform (DWFT) are time-frequency analysis techniques [20], which is adequate for analysing the translation invariants of images. Because these multiple directional maps, namely the vertical, horizontal and diagonal orientation maps, are of exactly equal size with the input image, thus they can be are directly normalized and jointly fused into a global orientation feature map. It is convenient to be able to extract directional patches within the global orientation feature map. In our model, directional patches with preceding K (e.g. K = 15) maximum values are selected from the global orientation feature map. Fig. 1 (b) display some detected directional patches from the global orientation feature map for locating the salient objects in the input images. As shown in Fig. 4 (b), the designed procedure is capable of extracting the directional patches in the salient objects. In [2], the centroid of the detected directional patches (the white cross as is illustrated in Fig. 1 (c)) is used to represent the spatial center of the salient object.

In this paper, in consideration of the sparseness of directional patches as well as each patch probably having a distinct weight during the estimating of the centroid, a weighted centroid calculation algorithm is devised to predict the spatial center of the salient object more accurately:

$$Centr\left(x',y'\right)=\left(\frac{\sum_{i=1}^{K}x_i DP_i(x_i,y_i)\omega_i}{\sum_{i=1}^{K}DP_i(x_i,y_i)\omega_i},\frac{\sum_{i=1}^{K}y_i DP_i(x_i,y_i)\omega_i}{\sum_{i=1}^{K}DP_i(x_i,y_i)\omega_i}\right);\quad (1)$$

$$\omega_i=\frac{\underset{j=1,j\neq i}{\arg\min}\left\{\left\|DP_i\text{ - }DP_j\right\|_2\right\}}{\sum_{i=1}^{K}\underset{j=1,j\neq i}{\arg\min}\left\{\left\|DP_i\text{ - }DP_j\right\|_2\right\}},s.t.\ \sum_{i=1}^{K}\omega_i=1.\quad (2)$$

where DP_i (i=1, 2, …, K) denote the K detected directional patches, and $\|\ \|_2$ represents Euclidean distance between two directional patches.

The aim of the weighted centroid computation algorithm is to depress the effect of these densely clustered patches, then utilizing the Euclidean distance to limit excessive intensive directional patches for estimating the center of salient object more accurately. Therefore, we can utilize the weighted centroid to represent the spatial center of the salient object. Fig. 1 (d) show some typical examples for fixing the salient objects of the images, it can be noted that the weighted centroid procedure is able to predict the center of salient object more precisely (which are closer to the ground truths, see Fig. 1 (e)) than the conventional method [2].

Then, assume that $posi(x,y)$ represents the spatial position with pixel coordinates (x,y) in the original image and spatial position center prior can be defined:

$$Cenpri(x,y)=\exp(\frac{-\left\|pos(x,y)-Centr\right\|_2}{2\sigma^2}),\quad (3)$$

where σ^2 adjusts the strength of spatial weighting.

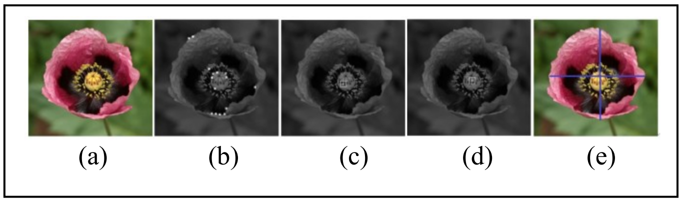

Fig. 1. Example of different center prior maps. (a) Input image; (b) the detected directional patches; (c) the centroid of the detected directional patches [2]; (d) the devised weighted centroid caculation algorithm; (e) Ground truths.

B. Colour contrast prior

Colour is an essential attribute of object; therefore, colour contrast feature is applied as intrinsic and forward visual cue to acquire the contrast information of the object for saliency detection.

Above all, the input image is processed by Simple Linear Iterative Clustering (SLIC) algorithm [21] to produce superpixels, and each superpixel composes of a feature vector $V\left[L,a,b,x,y\right]$, where (L,a,b) represents the colour information of the superpixel in the image, and therein (x, y) denotes the spatial coordinates. With regard to an individual superpixel i, the colour mean value c_i in CIELAB colour space, the average position of the superpixel's coordinates p_i, then the spatial weighted colour contrast of superpixel i for arbitrary superpixel j can be established [22]:

$$S_p\left(i\right)=\sum_{j\neq i}\left\|c_i-c_j\right\|\cdot\exp\left(-\frac{\left\|p_i-p_j\right\|_2}{2\theta^2}\right),\quad (4)$$

where θ^2 controls the spatial magnitude.

C. Sparse background feature

In the light of image content understanding, both the spatial position center prior and colour contrast prior can be considered as forward visual features for salient object detection. In this section, we attempt to incorporate a 'backward' visual cue for saliency detection. The original idea is that a reliable 'backward' visual cue will be propitious to tackle with the foreground and background of the image; meanwhile, it can also benefit of eliminating the background noises during saliency estimation.

In general, the surrounding regions around the image borders can provide a dependable background feature for saliency detection [23-27]. To obtain better background features, the SLIC method [21] is used to segment the image into uniform and compact superpixels. The superpixels around the image borders are adopted and extracted to construct the background template for dictionary learning [22-24], which is represented by $\mathrm{H} = \left[h_1, h_2, ..., h_m\right]$, in which m is the number of boundary superpixels. By using the constructed background template H as a dictionary of sparse, the sparse coefficient α_i of superpixel i can be solved as follows [22-24]:

$$\alpha_i = \arg\min_{\alpha_i} \sum_{i=1}^{m} \left\| x_i - H\alpha_i \right\|_2 + \lambda \sum_{i=1}^{m} \left\| \alpha_i \right\|_1 \ , \tag{5}$$

where x_i is the superpixel i around the image borders, and λ is a parameter larger than 0.

D. *Final Saliency Fusion*

Lastly, in order to take the merits of both the two types of the 'forward' and 'backward' visual cues for saliency detection, the spatial position center prior map, colour contrast map, and the sparse background map are jointly combined to produce the final hybrid saliency map:

$$Salmap = S_p.*Cenpri + S_b.*Cenpri \ , \tag{6}$$

where .* denotes the pixel-by-pixel mathematical multiplication operator.

III. Experimental Results

For the sake of verifying the performance of the designed saliency-detection system, the publicly available MSAR datasets [10] were tested for evaluation and comparison. In addition, seven typical state-of-the-art saliency-detection methods are chosen, containing the Frequency Tuned (FT) [28], Spectral Residual (SR) [5], Global Contrast (GC) [14], Robust Background Detection (RBD) [25], Geodesic Saliency (GS) [26], Background and Foreground Seed (BFS) [27], Dense and Sparse Reconstruction (DSR) [24]. In our experiments, the number of directional patches is with K=15.

In order to objectively compare these state-of-the-art methods with our proposed saliency-detection framework, three widely used criteria, namely the average precision (Pre), recall (Rec), and F-measure are counted quantitatively as well as statistically analysed. The F-measure is a comprehensive and overall indicator to weight precision more than recall.

Fig. 2 illustrates some experimental results of our proposed algorithm and individual state-of-the-art methods performed on the MSRA10K database. From Fig. 2 (i), it can be seen that our designed model can produce much more accurate saliency maps with the prominent objects highlighted. Meanwhile, compared with other existing saliency-detection models, the proposed scheme is unsusceptible to background noises and the boundaries of the salient objects are clearly restored.

In addition to visual analysis of different models, we also compared the performance of the proposed method with other seven state-of-the-art saliency-detection methods. Fig. 3 shows the precision, recall and the F-measure values of all the different methods. From comparisons, we can note that the average precision (*Pre*) of RBD, GS, BFS, DSR and our devised models exceed 70%. And the proposed method achieves the greatest overall detection accuracy in term of the precision, recall rate and F-measure. The extensive experimental results show that our proposed model is efficient and outperforms the other existing saliency-detection methods.

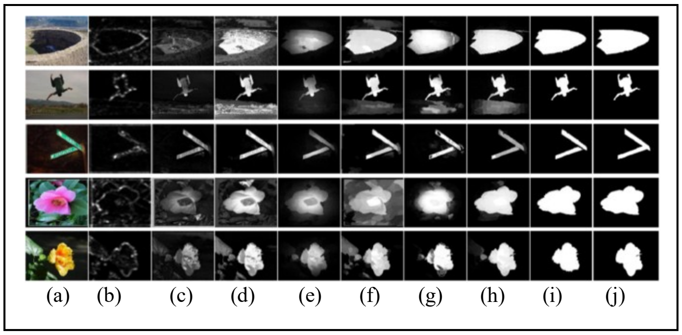

Fig. 2. Comparison of different state-of-the-art saliency detection models based on the MSRA database. (a) Input images; (b) SR; (c) FT; (d) GC; (e) RBD; (f) GS; (g) BFS; (h) DSR]; (i) The proposed method; (j) Ground truths.

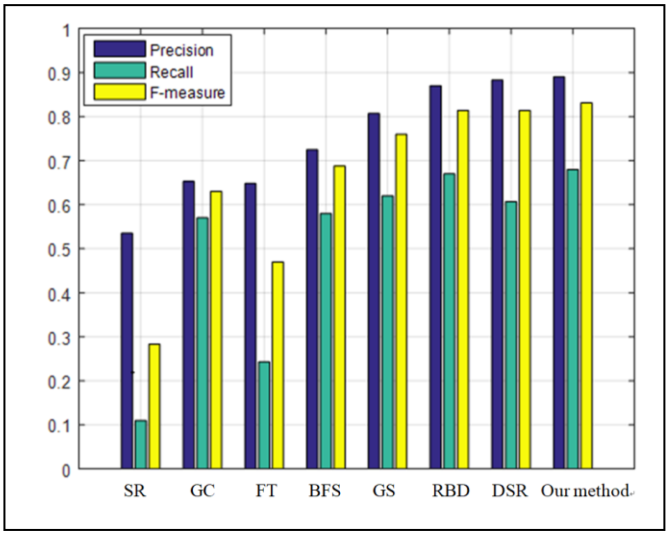

Fig. 3. Comparison of different saliency-detection methods in terms of average precision, recall, and F-measure based on the MSRA dataset.

IV. Conclusion and Discussion

In this paper, we have designed a visual saliency detection model based on spatial position prior, cues and features. To determine the spatial position of the salient reliably, we design a weighted centroid estimation algorithm with aid of DWDT. Furthermore, by integrating the forward colour contrast and the backward sparse background visual cues, a final compounded saliency map can be generated. We have performed testing of our method on three publicly available datasets, and experiments show that the proposed model is able to producing satisfactory and promising results compared with the state-of-the-art methods.

References

[1] M. Jian, Q. Qi, J. Dong, Y. Yin, K. M. Lam, Integrating QDWD with Pattern Distinctness and Local Contrast for Underwater Saliency Detection, Journal of Visual Communication and Image Representation, Vol. 53, pp. 31–41, 2018.

[2] M. Jian, W. Zhang, H. Yu, et al. Saliency detection based on directional patches extraction and principal local color contrast. Journal of Visual Communication and Image Representation, 2018, 57: 1-11.

[3] 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, 1998, 20 (11): 1254-1259.

[4] J. Harel, C. Koch and P. Perona. Graph-based visual saliency. Advances in neural information processing systems, 2006: 545-552

[5] X. Hou and L. Zhang. Saliency detection: a spectral residual approach. IEEE CVPR, 2007: 1-8.

[6] C. Guo, Q. Ma, L. Zhang. Spatio-temporal Saliency detection using phase spectrum of quaternion fourier transform. IEEE CVPR, 2008: 1-8.

[7] R. Achanta, S. Hemami, Q. Wang, J. Wan, Y. Yuan, Deep metric learning for crowdedness regression, IEEE Trans. Circuits System and Video Technology, https://10.1109/TCSVT.2017.2703920.

[8] R. Achanta and S. Süsstrunk. Saliency Detection for Content-aware Image Resizing. in IEEE International Conference on Image Processing, 2009.

[9] E. Rahtu, J. Kannala, M. Salo, et al. Segmenting salient objects from images and videos. In Proc. 11th European Conference on Computer Vision, 2010: 366–379.

[10] T. Liu, Z. Yuan, J. Sun, et al. Learning to detect a salient object. IEEE CVPR, 2011, 33: 353-367.

[11] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-Aware Saliency Detection. IEEE Trans. on Pattern Analysis and Machine Intelligence, Oct. 2012, 34 (10): 1915-1926.

[12] Y. Xie, H. Lu, M. Yang. Bayesian Saliency via Low and Mid Level Cues. IEEE Transaction on Image Processing, 2013.

[13] C. Yang, L. Zhang, H. Lu, M. Yang. Saliency Detection via Graph-Based Manifold Ranking. CVPR 2013.

[14] M. Cheng, N. Mitra, X. Huang, et al. Global contrast based salient region detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37 (3): 569–582.

[15] N. Liu, J. Han, Dhsnet: Deep hierarchical saliency network for salient object detection. IEEE CVPR, 2016: 678-686.

[16] A. Oliva, A. Torralba,M. Castelhano, and J. Henderson. Top down control of visual attention in object detection. In ICIP, volume 1, 2003, 253–256.

[17] C. Kanan, M. Tong, L. Zhang, and G. Cottrell. Sun: Top down saliency using natural statistics. Visual Cognition, 2009 17(6-7):979–1003.

[18] H. Cholakkal, J. Johnson and D. Rajan. A classifier-guided approach for top-down salient object detection. Signal Processing: Image Communication, 2016, 45: 24-40..

[19] M. Jian, K. M. Lam, J. Dong, L. Shen, "Visual-patch-attention-aware Saliency Detection", IEEE Transactions on Cybernetics, Vol. 45, No. 8, pp. 1575-1586, 2015.

[20] M. Unser, Texture classification and segmentation using wavelet frames, IEEE Trans. Image Process. 4 (11) (1995) 1549–1560.

[21] R. Achanta, A. Shaji, K. Smith, et al., SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2274-2282.

[22] C. Yang, L. Zhang and H. Lu. Graph-regularized saliency detection with convex-hull-based center prior. IEEE Signal Processing Letters, 2013, 20 (7): 637-640.

[23] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, “Saliency detection via graph-based manifold ranking,” in CVPR, 2013.

[24] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang. Saliency detection via dense and sparse reconstruction. In ICCV, 2013, 2976–2983.

[25] W. Zhu, S. Liang, Y. Wei, et al. Saliency optimization from robust background detection. IEEE CVPR, 2014: 2814-2821.

[26] Y. Wei, F. Wen, W. Zhu, and J. Sun. Geodesic saliency using background priors. In ECCV, 2012, vol. 7574, pp. 29–42.

[27] J. Wang, H. Lu, X. Li, et al. Saliency detection via background and foreground seed selection. Neurocomputing, 2015, 152: 359-368.

[28] R. Achanta, S. Hemami, F. Estrada, et al. Frequency-Tuned Salient Region Detection. IEEE CVPR, 2009: 1597-1604.