

Saliency detection via combining global shape and local cue estimation

*Qiang Qi*¹, *Muwei Jian*^{1, 2*}, *Yilong Yin*¹, *Junyu Dong*², *Wenyin Zhang*³, *Hui Yu*⁴

¹School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan, China

²Department of Computer Science and Technology, Ocean University of China, Qingdao, China

³School of Information Science and Engineering, Linyi University, Linyi, China

⁴School of Creative Technologies, University of Portsmouth, Portsmouth, UK

*(Correspondence author) E-mail: jianmuweihk@163.com

Abstract. Recently, saliency detection has become a hot issue in computer vision. In this paper, a novel framework for image saliency detection is introduced by modeling global shape and local cue estimation simultaneously. Firstly, Quaternionic Distance Based Weber Descriptor (QDWD), which was initially designed for detecting outliers in color images, is used to model the salient object shape in an image. Secondly, we detect local saliency based on the reconstruction error by using a locality-constrained linear coding algorithm. Finally, by integrating global shape with local cue, a reliable saliency map can be computed and estimated. Experimental results, based on two widely used and openly available databases, show that the proposed method can produce reliable and promising results, compared to other state-of-the-art saliency-detection algorithms.

Keywords: Saliency detection; QDWD; Locality-constrained linear coding, Local cue

1 Introduction

Visual saliency detection is used to predict the most interesting regions or objects in an image, which is extremely valuable in many various computer-vision applications, such as video surveillance [1], vision tracking [2, 3], image classification/retrieval [4], image retargeting [5], image segmentation [6], etc.

Saliency detection models can be categorized from different perspectives. Basically, most approaches adopt a bottom-up approach via low-level features while a few works incorporate a top-down solution by task driven. In the pioneering work [7] by Itti et al., a center-surround contrast method, based on the multi-scale image features, was presented for saliency detection. Later, a large number of methods were proposed, including the graph-based saliency model, proposed by Harel et al. [8] and

the fuzzy growing contrast method, proposed by Ma and Zhang [9]. In [10], a salient object segmentation method by using a saliency measure and a conditional random field (CRF) model was proposed. The works in [11], a new saliency model, which is based on the image regions and context-aware, was presented for saliency detection. Borji et al. [12] made an exhaustive comparison of thirty-five saliency models, based on fifty-four challenging synthetic patterns, three image datasets, and two video datasets. In [13], the final saliency maps are obtained by computing the covariance matrices of image features. Yang et al. [14] used the contrast, center and smoothness priors to estimate the saliency maps. The works in [15], a salient object detection algorithm was designed in multiple scales, based on the superpixels. Cheng et al. [16] proposed a novel saliency framework based on regional contrast (RC) to compute the global-contrast differences and spatial-weighted-coherence scores. In [17], a novel model based on Cellular Automata (CA) was proposed to compute the saliency of the objects. Jian et al. [18] proposed a visual-attention-aware model to mimic the human visual system (HVS) for salient-object detection. In [19], a bottom-up saliency-detection method was proposed by integrating directional, center and color cues. In [20], a novel self-paced multiple-instance learning (SP-MIL) framework, which considers both multiple instance learning (MIL) and self-paced learning (SPL), was proposed for co-saliency detection. The works in [21], a multistage saliency detection framework based on multilayer cellular automata (MCA) was proposed. Lu et al. [22] proposed a saliency detection framework, which exploring the fusion of various saliency models in a manner of bootstrap learning. In [23], a bottom-up saliency model that both consider the background and foreground cues was proposed.

The top-down methods need to consider both visual information and prior knowledge, thus these models usually contain complex learning process. In [24], a top-down saliency model, which using the global scene configuration, was proposed for saliency detection. Cholakkal et al. [25] adopted three locality constraints to compute the final saliency maps. In [26], a novel top-down saliency model that jointly learns a Conditional Random Field (CRF) and a visual dictionary was proposed. The works in [27] used a few exemplars and deep association to estimate the saliency of the image. In [28], a top-down contextual weighting saliency model, which incorporates high-level knowledge of the gist context of images was proposed.

In this paper, a novel saliency-detection method by modeling global shape and local estimation, is proposed for saliency detection. Different from existing saliency detection models, our proposed method integrates both global information and local features into a unified framework. We first computed the QDWD, which was initially designed for detecting outliers in color images, to represent the salient object shape in an image. Then, we incorporated an improved locality-constrained linear coding algorithm (LLC) to detect the local saliency. Finally, the two maps are combined to represent the most important saliency. In order to evaluate the performance of our proposed method, we carried out some experiments on two datasets and the comparison results with other state-of-the-art saliency-detection algorithms show that our approach is effective and efficient for saliency detection.

The rest of the paper is organized as follows. In Section 2, we introduce the proposed saliency-detection algorithm in detail. In section 3, we demonstrate our experimental results based on two image datasets and compared the results with other eight saliency-detection methods. The paper closes with a conclusion in Section 4.

2 The proposed saliency detection model

In this section, we present the proposed method, which incorporates the global shape of the salient object by using QDWD and local cue by utilizing a locality-constrained method, to simulate saliency detection. QDWD will first be described, followed by the locality-constrained method. All these different types of information are fused to form a saliency map, which indicates the effective fusion of global and local information as well.

2.1 Quaternionic Distance Based Weber Descriptor

The global methods [16, 29, 30], which are characterized by holistic rarity and uniqueness, take the entire image information into consideration to detect the salient regions, and thus detect large objects and uniformly assign saliency values to the contained regions. Unlike local methods, which are sensitive to high frequency image contents like edges and noise, global methods are less effective when the textured regions of salient objects are similar to the background. In order to reduce the effect of background and generate saliency maps with little noise, QDWD, which was initially proposed to detect the outliers and edges in an image [31], is utilized to represent the global shape information for the HVS to detect saliency [19, 35].

A quaternion \dot{q} is made up of one real part and three imaginary parts, as follows:

$$\dot{q} = a + ib + jc + kd, \quad (1)$$

where $a, b, c, d \in \mathfrak{R}$; i, j, k are complex operators; a is the real part; and $\{ib, jc, kd\}$ are the imaginary parts. By considering two pixels in a color image, $\dot{q}_m = r_m i + g_m j + b_m k$ and $\dot{q}_n = r_n i + g_n j + b_n k$, different types of distances can be defined to measure the distance of quaternions. Let $D_t(\dot{q}_m, \dot{q}_n)$ represents the t^{th} type Quaternionic distances (QD) of quaternions \dot{q}_m and \dot{q}_n , we compute the different types of quaternions distances $D_t(\dot{q}_m, \dot{q}_n)$ using the method in [31]. The computed quaternionic distances, which can be defined as the increments between two quaternions, can also be used to measure the similarity between different quaternions from different perspective viewpoints. That is to say, the quaternionic increment between quaternions can be computed by using quaternionic distances.

Assume that \dot{q}_c denotes the center quaternion in a local patch, and \dot{q}_l ($l \in L$, where L is the index set) represents the residual quaternions in the patch. Thus, the total quaternionic increment in a local patch can be written as $\sum_{l=0}^{l=7} D_t(\dot{q}_c, \dot{q}_l)$. With the aid of $|\dot{q}_c|$ as the quaternionic intensity, the differential features of QDWD defined by D_t can be represented as follows:

$$\Phi_2'(\dot{q}_c) = \arctan\left(\frac{\sum_{l \in L} D_t(\dot{q}_c, \dot{q}_l)}{|\dot{q}_c|}\right). \quad (2)$$

The nonlinear mapping, $\arctan(\cdot)$, aims to enhance $\Phi'_2(\dot{q}_c)$ to become more robust.

In order to achieve a better performance, we normalize these differential features achieved from Eq. (2) and add them to form an integrated global shape map. Thus, the global saliency can be defined as follows:

$$G = \frac{1}{N} \sum_{t=1}^N \Phi'_2(\dot{q}_c) \quad (3)$$

where $\Phi'_2(\dot{q}_c)$ is the differential features defined in Eq. (2), and $N=6$ in our implement.

Fig. 1(a) shows an example of the QDWD features for an input color image. Fig. 1 (b)-(f) are the differential feature maps $\Phi'_2(\dot{q}_c)$ of QDWD, obtained by using $D_t (t=1,2,3,4,6)$. The image generated by the 5th QD is not given, since it is equivalent to the 3rd QD. As shown in Fig. 1 (b)-(f), these QDWD features can be utilized to reflect the global shape information of the salient object. Then, we normalize these global shape maps and add them to form an integrated holistic global shape map G . Fig. 1(g) and Fig. 2(b) show the fusion of the different feature maps to form an integrated global shape map, which can be utilized for saliency detection. More details about QDWD features for saliency detection, please refer to [19].

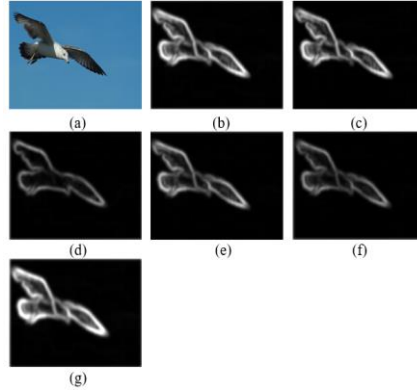


Fig. 1 An example of the QDWD features. (a) an input image, (b)-(f) the differential feature maps of QDWD produced by different QDs, and (g) the integrated global shape map.

2.2 Locality-Constrained Method

The motivation of local estimation is the local outliers, which are standing out from their neighbors with different colors or textures and tend to attract human attention. In order to detect local outliers and get acceptable performance, local coordinate coding method, which described the locality is more essential than sparsity, has been used in saliency detection. Furthermore, the proposed QDWD model in Section 2.1 only takes the global shape information into consideration, which highlight the edges of salient

object but miss some local cues. Thus, we employ an approximated algorithm based on locality-constrained linear coding (LLC) [33] to estimate the local saliency.

For a given image, we first over-segmented the image into N regions, $\{r_i\}, i=1,2,\dots,N$. For each region r_i , let \mathbf{X} be a set of 64-dimensions local descriptors extracted from the image, and $\mathbf{X} = [x_i^0, x_i^1, \dots, x_i^{63}]^T, i=1,2,\dots,N$. Therefore, the original function of LLC method is written as follows [33]:

$$\min_{\mathbf{B}} \left\{ \sum_{i=1}^N (\|x_i - \mathbf{D}b_i\|^2 + \omega \|dr_i .* b_i\|^2) \right\}, \quad s.t. \quad \mathbf{1}^T b_i = 1, \quad \forall_i, \quad (4)$$

where $\mathbf{B} = [b_1, b_2, \dots, b_N]$ is the set of codes for \mathbf{X} , $\mathbf{D} = [d_1, d_2, \dots, d_M]$ is the codebook with M entries, and the parameter ω is used to balance the weight between the penalty term and regularization term. The constraint $\mathbf{1}^T b_i = 1$ follows the shift-invariant requirements of the LLC coding and $.*$ denotes an element-wise multiplication. Here, the dr_i is the locality adaptor that gives different freedom for each codebook vector based on its similarity to the input descriptor x_i , and is defined:

$$dr_i = \exp\left(\frac{\text{dist}(x_i, \mathbf{D})}{\lambda}\right), \quad (5)$$

where $\text{dist}(x_i, \mathbf{D}) = [\text{dist}(x_i, d_1), \text{dist}(x_i, d_2), \dots, \text{dist}(x_i, d_M)]^T$, $\text{dist}(x_i, d_i)$ denotes the Euclidean distance between x_i and the codebook vector d_i , λ is used to adjust the weight decay speed for the locality adaptor, and M is the number of elements in the codebook. More details about LLC can be referred to [32, 33].

In this paper, we adopted an approximated LLC algorithm [32] to detect the local cue. We consider the K nearest neighbors in spatial as the local basis \mathbf{D}_i owing to the vector b_i in Eq. (4) with a few non-zero values, which means that it is sparse in some extent. It should be noted that the K is smaller than the size of the original codebook M . Thus, the Eq. (4) can be rewritten as follows:

$$\min_{\mathbf{B}} \left(\sum_{i=1}^N \|x_i - \mathbf{D}_i b_i\|^2 \right), \quad s.t. \quad \mathbf{1}^T b_i = 1, \quad \forall_i, \quad (6)$$

where \mathbf{D}_i denotes the new codebook for each region $r_i, i=1,2,\dots,N$ and K is the size of the new codebook and is empirically set at $K = 2M/3$.

Unlike the traditional LLC algorithm, solving the approximated LLC algorithm [32] is simple and the solution can be derived analytically by

$$b_i = 1/(\mathbf{C}_i + \omega \times \text{dig}(\mathbf{C}_i)), \quad (7)$$

$$\square \quad b_i = b_i / \mathbf{1}^T b_i, \quad (8)$$

where $\mathbf{C}_i = (\mathbf{D}_i - \mathbf{1}x_i^T)(\mathbf{D}_i - \mathbf{1}x_i^T)^T$ represents the covariance matrix of the feature and ω is a regularization parameter, which is set to be 0.1 in the proposed algorithm. As

the solution of the improved LLC method is simple and fast, therefore, we defined the local saliency value of the region r_i as follows [32]:

$$S(r_i) = \|x_i - D_i \hat{b}_i\|^2, \quad (9)$$

where \hat{b}_i is the solution of Eq. (6), which is achieved by Eq. (7) and Eq. (8).

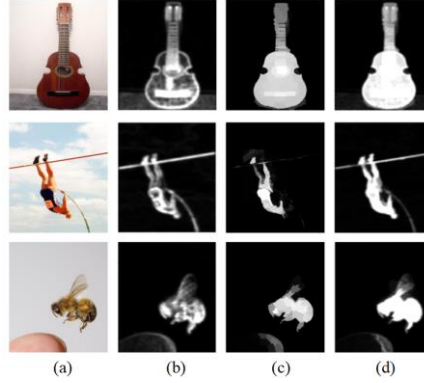


Fig. 2. The proposed scheme for saliency detection. (a) input images, (b) global shape saliency maps, (c) local saliency maps, (d) the final saliency maps.

2.3 Final Saliency Fusion

We have already generated global shape saliency map G and local saliency map S for an input image. As is shown in Fig. 2 (b), global shape maps detect salient object accurately with more complete shape due to the QDWD, which highlight the edges of salient object but miss some inside object cues. While the local saliency maps achieve more reliable local details owing to the locality-constrained coding model. By integrating the two kind of maps, both global and local information are take into account simultaneously. Therefore, the final saliency map can be defined as follows:

$$Sal = G \oplus S, \quad (10)$$

Where \oplus denotes the addition operator in Eq. (10). It should be noted that we explore a lot of methods for integration of the global shape maps and local saliency maps, and we find that combining them straightforwardly can obtain the best saliency maps. Although the fusion algorithm seems simple, it achieves satisfactory performance as shown in Fig. 2(d).

3 Experimental results

We present some experimental results based on two widely used datasets: MSRA10K [16] and THUR15K [34], and compare our method with other eight state-of-the-art methods including the Itti [7], Graph-Based (GB) [8], Segmenting-E. Rahtu-J. Kannala (SEG) [10], Context-aware (CA) [11], Nonlinearly Covariance (NC) [13],

Graph-regularized (GR) [14], Multi-Scale Superpixel (MSS) [15] and Background-Single Cellular Automata (BSCA) [17]. In order to quantitatively compare the state-of-the-art saliency-detection methods, the average precision, recall, and F -measure are utilized to measure the quality of the saliency maps.

3.1 MSRA10K

We first evaluate the proposed algorithm on the MSRA10K database. It contains 10,000 images with their pixel-wise ground truth, which are randomly chosen from the MSRA dataset.

Fig. 3 shows some saliency detection results of different methods based on MSRA10K database. It can be seen that the proposed algorithm consistently generates saliency maps closer to the ground truth. We also compared the performance of the proposed method with other eight state-of-the-art saliency detection methods. Fig. 4 shows the precision, recall and the F -measure values of all the different methods. From the comparison, we can see that our proposed method outperforms the other state-of-the-art methods. Meanwhile, our proposed method outperforms the GR [8] method which only used the local contrast priors.

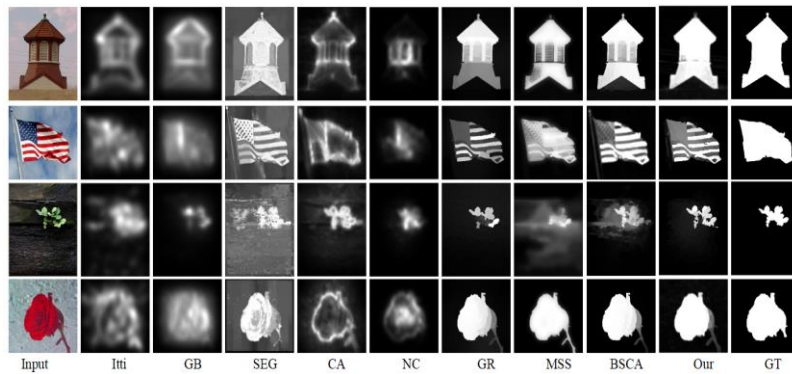


Fig. 3 Saliency detection results of different methods based on MSRA10K database.

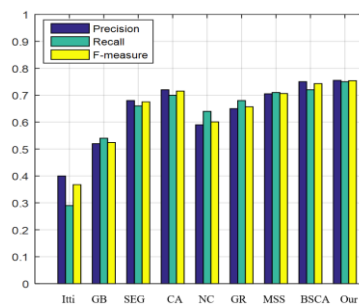


Fig. 4 Comparison of different saliency-detection methods in terms of average precision, recall, and F -measure based on the MSRA10K database.

3.2 THUR15K

We also test the proposed model on the THUR15K database. Fig. 5 shows some results of saliency maps generated by nine different algorithms including our method. As shown in Fig. 5, the proposed model produces saliency maps with better outlines than other eight methods attributing to the contribution of QDWD, which is used to detect the object’s global shape information.

For objective evaluation, we also used the precision, recall and the F -measure to assess the performance of the nine different methods. Fig. 6 shows the comparisons of different methods according to different evaluation criteria. From Fig. 6, the proposed method achieves the best performance on the THUR15K database in term of both precision and F -measure values.

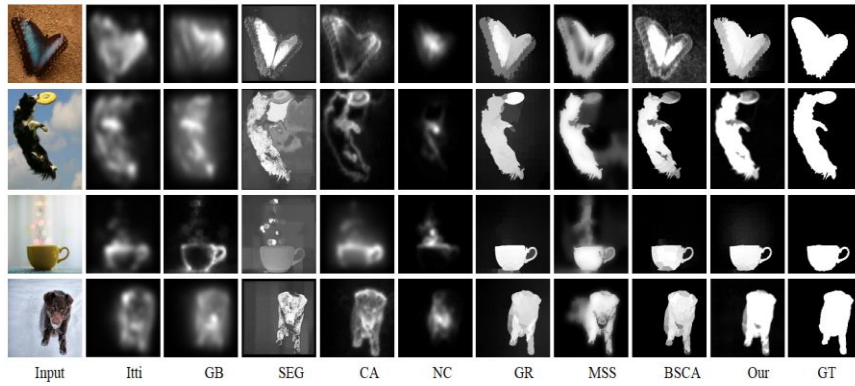


Fig. 5 Qualitative comparisons of different approaches based on THUR15K database.

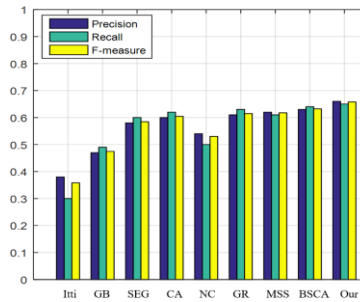


Fig. 6 Comparison of different saliency-detection methods in terms of average precision, recall, and F -measure based on THUR15K database.

4 Conclusion and discussion

This paper proposes a novel bottom-up method, which incorporates the global shape and local estimation into a unified framework for saliency detection. The global shape maps are obtained by using QDWD rather than directly computing global contrast. The local saliency cues are achieved by utilizing a locality-constrained method. The final saliency maps are generated by integrating the global with the local saliency maps, which can take both global and local cues into account simultaneously. Experiments show that the proposed algorithm can produce satisfactory results.

Acknowledgments

This work was supported by National Natural Science Foundation of China (NSFC) (61601427); Natural Science Foundation of Shandong Province (ZR2015FQ011); Applied Basic Research Project of Qingdao (16-5-1-4-jch); China Postdoctoral Science Foundation funded project (2016M590659); Postdoctoral Science Foundation of Shandong Province (201603045); Qingdao Postdoctoral Science Foundation funded project (861605040008) and The Fundamental Research Funds for the Central Universities (201511008, 30020084851).

References

1. Huang, K., et al.: Biologically inspired features for scene classification in video surveillance. *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 1, pp. 307 - 313. (2010)
2. Zhang, K., Zhang, L., Yang, M. H.: Fast compressive tracking. *IEEE Trans. PAMI*, vol. 36, no. 10, pp. 2002-2015. (2014)
3. Zhang, K., Liu, Q. and Wu, Y.: Robust Visual Tracking via Convolutional Networks Without Training. *IEEE TIP*, vol. 25, no. 4, pp. 1779-1792. (2016)
4. Jian, M.W., Dong, J.Y., and Ma, J.: Image retrieval using wavelet-based salient regions. *Imaging Sci. J.*, vol. 59, no. 4, pp. 219 - 231, Aug. (2011)
5. Zhu, G., Wang, Q., Yuan, Y., and Yan, P.: Learning saliency by MRF and differential threshold. *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 2032 - 2043, Dec. (2013)
6. Hsu, C.Y., and Ding, J.J.: Efficient image segmentation algorithm using SLIC superpixels and boundary-focused region merging. *Information, Communications and Signal Processing*, pp. 1-5. (2013)
7. Itti, L., Koch, C., Niebur, E.: A model of saliency based visual attention for rapid scene analysis. *IEEE Trans. PAMI*, vol. 20, no. 11, pp. 1254-1259. (1998)
8. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. *Advances in NIPS*, pp. 545-552. (2006)
9. Ma, Y.F., Zhang, H.J.: Contrast-based image attention analysis by using fuzzy growing. *ACM conference on Multimedia*, pp. 374-381. (2003)
10. Rahtu, E., Kannala, J., Salo, M., and Heikkilä, J.: Segmenting salient objects from images and videos. In *Proc. 11th ECCV*, pp. 366 - 379. (2010)

11. Goferman, S., Zelnik-Manor, L., Tal, A.: Context-aware saliency detection. *IEEE Transactions on PAMI*, vol. 34, no. 10, pp. 1915-1926. (2012)
12. Borji, A., Sihite, D.N., and Itti, L.: Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE TIP*, vol. 22, no. 1, pp. 55 - 69. (2013)
13. Erdem, E., Erdem, A.: Visual saliency estimation by nonlinearly integrating features using region covariances. *Journal of vision*, vol. 13, no. 4, pp. 11-11. (2013)
14. Yang, C., Zhang, L., Lu, H.: Graph-regularized saliency detection with convex-hull-based center prior. *IEEE Signal Proces. Let.*, vol. 20, no. 7, pp. 637-640. (2013)
15. Tong, N., Lu, H., Zhang, L., et al.: Saliency detection with multi-scale superpixels. *IEEE Signal Processing Letters*, vol. 21, no. 9, pp. 1035-1039. (2014)
16. Cheng, M.M., Mitra, N.J., Huang, X., et al.: Global contrast based salient region detection. *IEEE Trans. PAMI*, vol. 37, no. 3, pp. 569 - 582. (2015)
17. Qin, Y., Lu, H., Xu, Y., et al.: Saliency detection via cellular automata. In *IEEE CVPR*, pp. 110-119. (2015)
18. Jian, M., Lam, K.M., Dong, J.J., Shen, L.L.: Visual-Patch-Attention-Aware Saliency Detection. *IEEE T. Cybernetics* vol. 45, no. 8, pp. 1575-1586. (2015)
19. Jian, M.M., Qi, Q., Sun, Y., Lam, K.M., et al.: Saliency Detection Using Quaternionic Distance Based Weber Descriptor and Object Cues. *APSIPA ASC'2016*, Korean. (2016)
20. Zhang, D., Meng, D., Han, J.: Co-saliency detection via a self-paced multiple-instance learning framework. *IEEE transactions on pattern analysis and machine intelligence*. (2016)
21. Wang, A., Wang, M.: RGB-D Salient Object Detection via Minimum Barrier Distance Transform and Saliency Fusion. *IEEE Signal Processing Letters*. (2017)
22. Lu, H., Zhang, X., Qi, J., et al.: Co-Bootstrapping Saliency. *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 414-425. (2017)
23. Lin, X., Yan, Z., Jiang, L.: Saliency Detection via Foreground and Background Seeds. *International Conference on Information Science and Applications*. Springer, Singapore, pp. 145-154. (2017)
24. Oliva, A., Torralba, A., Castelano, M.S., et al: Top-down control of visual attention in object detection. *IEEE ICIP*, vol. 1. (2003)
25. Cholakkal, H., Rajan, D., Johnson, J.: Top-down saliency with locality-constrained contextual sparse coding. *BMVC*. (2015)
26. Yang, J., Yang, M. H.: Top-down visual saliency via joint crf and dictionary learning. *IEEE Transactions on PAMI*, vol. 39, no. 3, pp. 576-588. (2017)
27. He, S., Lau, R.W.H., Yang, Q.: Exemplar-driven top-down saliency detection via deep association. In *IEEE CVPR*. pp. 5723-5732. (2016)
28. Rahman, I., Hollitt, C., Zhang, M.: Contextual-based top-down saliency feature weighting for target detection. *Machine Vision and Applications*, vol. 27, no. 6, pp. 893-914. (2016)
29. Achanta, R., Hemami, S., Estrada, F., and Susstrunk, S.: Frequency-tuned salient region detection. In *IEEE CVPR*, pp. 1597 - 1604. (2009)
30. Perazzi, F., Krahenbuhl, P., Pritch, Y., and Hornung, A.: Saliency filters: Contrast based filtering for salient region detection. In *IEEE CVPR*, pp. 733 - 740. (2012)
31. Lan, R.S., Zhou, Y.C., Tang, Y.: Quaternionic Weber Local Descriptor of Color Images. *IEEE Trans. on Circuits and Systems for Video Technology*. (2015)
32. Tong N, Lu H, Zhang Y, et al. Salient object detection via global and local cues. *Pattern Recognition*, 48(10): 3258-3267. (2015)
33. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In *IEEE CVPR*, pp. 3360 - 3367. (2010)
34. Cheng, M.M., Mitra, N.J., Huang, X.: Salient shape: Group saliency in image collections. *The Visual Computer*, vol. 30, vol. 4, pp. 443-453. (2014)
35. Muwei Jian, Qiang Qi, et al. "Saliency Detection Using Quaternionic Distance Based Weber Local Descriptor and Level Priors", *Multimedia Tools and Applications* (2017). DOI: 10.1007/s11042-017-5032-z