Iterative Saliency via Dynamic Image Region Partitioning

Fangjun Wen

Abstract—A novel object level saliency model is proposed in this letter via iterating saliency classification difference on dynamic image region partitioning. First, the model proposed solved three problems which is caused by static background based methods. Then dynamic background is represented and computed on the input image via dynamic image partitioning. Unlike existing static background based methods, we calculate saliency difference based on dynamic background rather than static image region. This strategy makes the saliency result more precisely to the location of image objects. We apply two saliency classification difference on the dynamic background. Second, saliency classification difference is iterated on saliency maps generating by dynamic image region partitioning. This makes the saliency results more robust. To get a more robust result, the dynamic image partitioning is operated on an image in four directions (i.e., left to right, right to left, top to bottom, bottom to top). Third, the final saliency map is generated by combining four saliency maps based on four direction scanning. The four direction combination enables the proposed method to uniformly highlight the salient object and simultaneously suppress the background effectively. Extensive experiments on two large dataset demonstrate that the proposed method performs favorably against the classic methods in terms of accuracy and efficiency.

***Index Terms*—**Dynamic image region partitioning, iterating, dynamic background, four direction scanning, saliency map.

I. Introduction

SALIENT object detection is the process of identifying the most distinguishable object that human vision system interested in a complex scene and a key attentional mechanism related to the basic survival skills. At the era of big mount of image and video data, it is needed to focus the finite computing resources to let computer have the ability to understand the surrounding environment as we humans do. Visual saliency plays an important role for intelligent computer vision and multimedia applications. The saliency-based applications can be normally classified into six categories nowadays, including retargeting [1], advertising [2], retrieval [3], summarization [4], compression [5] and recognition [6].

Saliency models can be generally categorized as two directions either task-independent bottom-up saliency or task-driven top-down saliency from research interests. Bottom-up saliency [7]-[18] is fast, involuntary, stimulus-driven and top-down saliency [20] is slower, voluntary and goal-driven. Borji et al. [21] notes that the two researching waves of saliency detection are eye fixation prediction [7]-[10] and salient object detection [11]-[18].

Researches from perceptual research [22-24] indicate that the most influential factor in low-level visual saliency is contrast. However, the calculation of contrast in previous works is based on various different types of image features, including color variation of individual pixels, edges and gradients, spatial frequencies, structure and distribution of image patches, histograms, multi-scale descriptors, or combinations thereof [15]. The definition of saliency difference is often being calculated only once and most of them all rely on complex classification models to detect the salient image region. Recently, methods that exploits background priors are proposed for saliency detection [17]-[19]. They model saliency detection as a ranking problem. By using the image patches on the image boundary as queries, the remaining nodes are ranked based on their relevance to the given query, which usually yields wrong ranking results when object taking place on image boundary, color feature of large area on the image boundary is almost the same as salient object or original image with 30 pixels pure black rectangle on image border.

Considering all the above mentioned issues, we propose a bottom-up saliency detection model based on following properties:

- Structure. To highlight salient pixels uniformly and efficiently, we exploiting image structure for saliency detection via superpixels.
- Region contrast. Local region contrast provides more visual information than pixels-based contrast.
- Iterative. Iterating ranking results exaggerates the ranking difference of superpixels.
- Dynamic image region partitioning. Dynamic image region partitioning makes the dynamic background possible. In contrast to static background-based methods, dynamic background yields more precisely saliency detection results.
- Four direction scanning. Four direction scanning makes the proposed saliency model more robust.

Based on these properties, this letter proposes a salient object detection model based on iterating saliency classification difference, dynamic image partitioning and four direction scanning.

The most related works are [17]-[19]. They are all static background-based methods and doing saliency classification by a complex mathematical model. Unlike their works, we have four main contributions as follows.

- First, dynamic background is generated to replace the static background by exploiting dynamic image region partitioning.
- Second, iterating saliency classification difference gradually exaggerates the saliency difference among superpixels so that salient object and image background can separate automatically finally.
- Third, four direction scanning makes the final saliency detection result more robust, which highlights the salient object and suppressing the background uniformly.
- Fourth, the saliency classification strategy that we choose is rather simple for that we only use Euclidean distance as the measurement approach.

We use the Precision and Recall (P-R) curve and Mean Absolute Error (MAE) to evaluate the proposed algorithm and 11 state-of-art methods on two benchmark datasets. Fig. 1 shows samples of saliency maps generated by state-of-art methods and our methods. Both quantitatively and qualitatively experimental results demonstrate that our algorithm performs favorably among all the evaluated methods, which bear out the validity of the principles used in proposed saliency model.

Fig. 1. Saliency maps (from left to right): input, ground truth, ISDIP, AC [11], FT [12], RC11 [13], LRMR [14], SF [15], GMR [17], BSCA [18]. ISDIP is the proposed algorithm without optimization. The proposed algorithm highlights the salient object uniformly.

Fig. 2. Dynamic Image Partitioning (from left to right): left to right dynamic image region partitioning, right to left dynamic image region partitioning, top to bottom dynamic image region partitioning, bottom to top dynamic image region partitioning.

II. Saliency via Dynamic Image Region Partitioning

The proposed approach is formulated based on superpixels [25] as they encode compact and structural information within a scene. As for static background based method, the final results will be affected by the position of salient object or color of background. Dynamic image partitioning can generate dynamic background which can solve three error detection caused by static background based methods. In contrast with the traditional static background based methods, we formulate the saliency detection problem as an iterating process which can finally separate the salient object and background by exaggerate the saliency difference among superpixels. With four direction combination, the saliency detection results is more robust than any single direction, two direction or three direction combination.

A. Dynamic Image Partitioning based Saliency Map

1) Image Features: Based on superpixels, we consider local region contrast, dynamic image partitioning, iterating saliency classification difference and four direction combination to compute saliency maps. An image is segmented into superpixels $\{c_i\}, i = 1, 2, ..., \text{N},$ where N is the number of segmented regions. We construct one combined feature vector for each superpixel c in the CIE RGB, CIE LAB and CIE XYZ color space. The feature vector of each superpixel c is defined as

$$f = (r_c, g_c, b_c, l_c, a_c, b_c, x_c, y_c, z_c), \tag{1}$$

2) Dynamic Background: Dynamic background is just a size adjustable image region with superpixels beginning with one side on image borderline. Lu [17] proposed a salient object detection method based on static background of superpixels on four sides of image border by using manifold ranking to ranking the superpixels. However, static background-based methods may cause three error detections when object taking place on image boundary, color of large area adjacent to the image boundary is almost the same as salient object or input image with 30 pixels pure black rectangle on image border. For the third issue, Lu [17] adopted an approach by doing a pre-processing on the input image through straightly moving 30 pixels when that exists. In our saliency model, there is no need to do any pre-processing to the input image.

3) Dynamic Image Region Partitioning: Based on a straight line which is parallel to each side of left, right, top or bottom borderline of the input image, we get a series of dynamic background which is the left side of the straight line and unclassified image region which is the right side of the straight line through moving the parallel straight line in a specific moving direction (i.e., left to right, right to left, top to bottom, bottom to top). Superpixels are named as either dynamic seeds which are inside dynamic background or unclassified seeds which are inside unclassified image region. By taking dynamic seeds as a whole, we use their average feature vector to rank all the unclassified seeds. Step-size of the moving line is initialized to a whole number which is one less than the square root of the total number of superpixels. We set r as the step-size.

We note the total number of pair of dynamic background and unclassified image region as number P.

Fig. 2 shows the four direction dynamic image region partitioning.

4) Saliency Measure: We measure the saliency difference between two nodes p and q by using Euclidean metric. In this letter, we set p as the average feature vector of dynamic seeds and q as the feature vector of each individual seed of unclassified seeds. Given a superpixel c_i and its neighboring regions, we define

$$p = \overline{\sum_{i=1}^{M} f_i}, \tag{2}$$

and

$$q = f_j,$$
$$j = 1 : N \;\&\; j \neq i, \tag{3}$$

M is the number of superpixels of dynamic seeds. We define the saliency measure of each individual seed of unclassified seeds as

$$s_j = \mathrm{D(p, q)}, \tag{4}$$

$\mathrm{D(p, q)}$ is the Euclidean distance between p and q. We note the saliency matrix of superpixels as W, where

$$W_j = \begin{cases} s_j & j = (M+1):N \\ 0 & j = 1:M, \end{cases} \tag{5}$$

5) Iterating Saliency Difference: Based on each pair of dynamic seeds and unclassified seeds, we get a saliency map S_k $(k = 1 : W/r$ or $H/r)$ based on the saliency matrix W_k calculated by the above defined saliency measure. The first saliency map is initialized as $S_1 = W_1$. From the second saliency map to the final saliency map, we set the iterative process of saliency matrix calculation as follows:

$$(S_{k+1})_j = (S_k)_j + D_j, \tag{6}$$

$$k = 1 : W/2r \text{ or } H/2r, \; j = 1 : N, \tag{7}$$

B. Four Direction Combination

The final saliency map is formed by combining the four saliency maps generated by four direction scanning including the left to right scanning based saliency map, the right to left scanning based saliency map, the top to bottom scanning based saliency map and the bottom to top scanning based saliency map.

After the process of dynamic image region partitioning, there is only one saliency map being generated. We note the left to right scanning based saliency map as S_L, right to left scanning based as S_R, top to bottom scanning based saliency map as S_T, bottom to top scanning based saliency map as S_B. The final saliency map S is defined as the linear combination of the four direction based saliency maps.

$$S = S_L \times S_R \times S_T \times S_B. \tag{8}$$

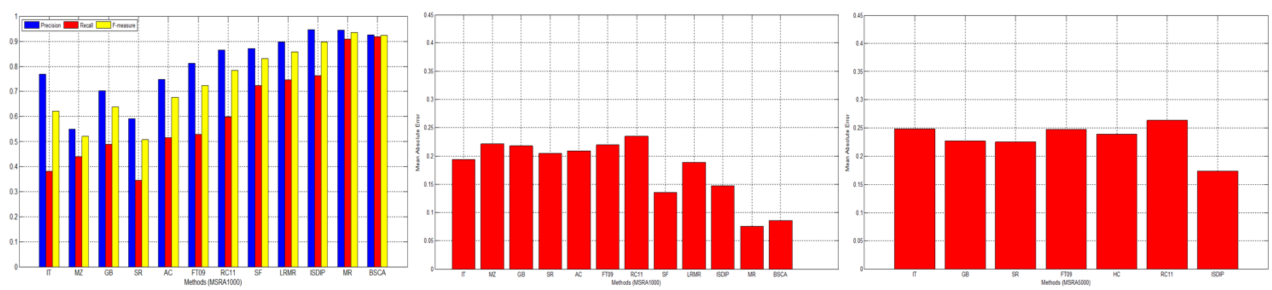

Fig. 2. The left most one figure is the F-measure curve on the ASD dataset and the right two figures are the MAE curves on the ASD dataset and MSRA dataset.

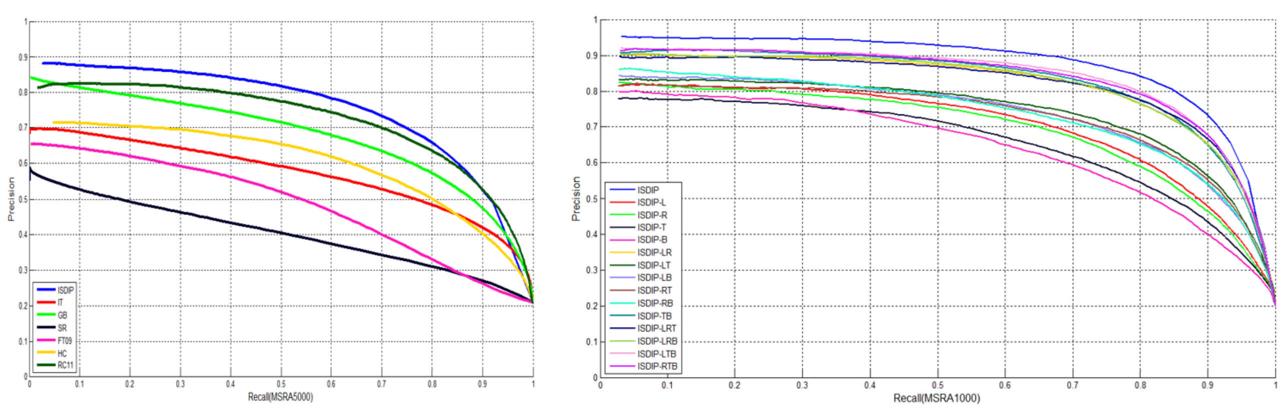

Fig. 3. The left figure is the Precision-Recall curve on the MSRA dataset and the right figure is the Precision-Recall curve of four direction combination on the ASD dataset.

III. EXPERIMENTS AND RESULTS

In this letter, we compare the proposed method with 11 state-of-art saliency detection approaches on two publicly available datasets to demonstrate its superiority. The ASD (MSRA1000) dataset is a salient object dataset of 1000 images selected from the MSRA dataset [20] with pixel-level ground truths [12]. MSRA dataset is a dataset with 5000 images with relatively complex scene and multi-objects in contrast to ASD dataset. For other algorithms, we use the implementations or the result maps provided by the authors for fair evaluation. All the experiments are run in the MATLAB 2011b platform on a PC with Intel i5-6500 CPU (3.2 GHz) and 4GB RAM. The 11 compared methods on the ASD dataset are: IT98 [7], MZ03 [8], GB07 [9], SR07 [10], AC08 [11], FT09 [12], RC11 [13], SF12 [15], LRMR12 [14], MR13 [17], BSCA16 [18].

We show comparative results with classic methods on the MSRA dataset since some other algorithms only provide saliency maps on the ASD dataset.

1) Experimental Setup:

In this letter, there exists only one external parameter which is the number of superpixel nodes and we set it with $N = 200$ in all experiments.

2) Saliency maps: Fig. 1 shows the comparison of the saliency maps generated by 8 methods including ours. The experiments show that the saliency maps generated by our methods can locate salient object more precisely than other methods and some other methods failed to locate the right salient object precisely especially for background based methods such as GMR [17] and BSCA [18]. What's more, the proposed algorithm solves three error detection cases caused by the background based saliency detection methods as mentioned in Part I. All the priorities show that the proposed method can both highlight the salient object uniformly and suppress the background effectively.

3) Quantitative evaluation: For a saliency map with intensity values in the range between [0, 255], we set the threshold from 0 to 255 with an increment of 5, obtaining 52 binary masks for each image. From Fig. 2 we could see that our proposed algorithm reaches the highest precision than other methods and the mean absolute error of the proposed algorithm is also lower than the classic methods on the ASD dataset. From Fig. 3 we could see that the proposed algorithm is favorably against the classic methods on the MSRA dataset and the proposed four direction combination is very useful for saliency map integration. From Fig. 4 we could see that the proposed algorithm reaches the highest precision when the dynamic background is set as half size of the input image. From Fig. 5 we could see that the four direction combination is rather effective.

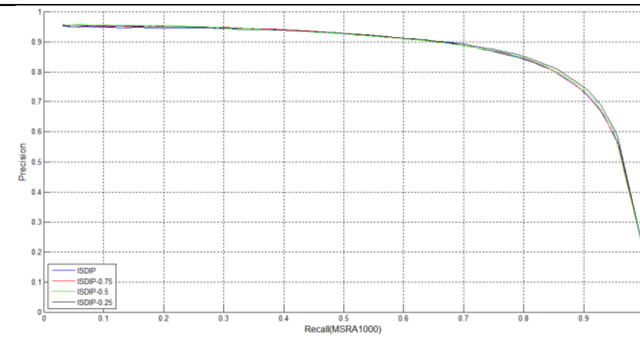

Fig. 4. Dynamic Background with different size. In this letter, we set the biggest dynamic background as half size of the input image on which we get the highest salient object detection precision.

IV. CONCLUSION AND DISCUSSION

In this letter, we present a bottom-up object level saliency detection model by exploiting dynamic image region partitioning, iterating saliency difference and four direction scanning. Firstly, we get more precise results than the background based methods with dynamic image region partitioning. Secondly, iterating the saliency classification difference automatically separates the salient object and the background. Thirdly, the four direction scanning makes our algorithm more robust. Saliency maps on a large public dataset demonstrate that the proposed method can highlight the whole object region uniformly and suppress the background region effectively. In addition, the proposed method performs favorably against the classic methods in accuracy, which shows that the proposed dynamic image region partitioning, iterating saliency classification results and four direction scanning are useful for saliency detection. In the future work, we will investigate a more complicated classification strategy to boost classification results at different scales and explore more applications of dynamic image region partitioning to other saliency algorithms.

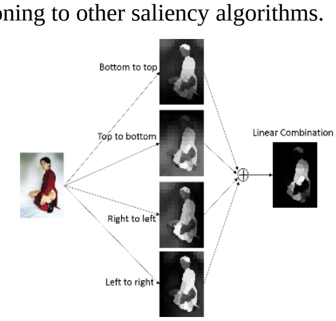

Fig. 5. The process of four direction combination of saliency maps.

REFERENCES

[1] Avidan, Shai, and A. Shamir. “Seam carving for content-aware image resizing.” Acm Transactions on Graphics 26.3(2007):10.

[2] Gao, W., Tian, Y., Huang, T., Yang, Q.: Vlogging: A survey of videoblogging technology on the web. ACM Computing Surveys 42(4), 15, 1–57 (2010).

[3] Fu, H., Chi, Z., Feng, D.: Attention-driven image interpretation with applicationto image retrieval. Pattern Recognition 39(9), 1604–1621 (2006).

[4] Ma, Y.F., Hua, X.S., Lu, L., Zhang, H.J.: A generic framework of user attention model and its application in video summarization. IEEE Transactions on Multimedia 7(5), 907–919 (2005).

[5] Li, Z., Qin, S., Itti, L.: Visual attention guided bit allocation in video compression.Image Vision Computing 29(1), 1–14 (2011).

[6] Parikh, D., Zitnick, C., Chen, T.: Determining patch saliency using low-level context.Berlin, Germany 2, 446–459 (2008).

[7] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” IEEE Trans. Patt. Anal. Mach. Intell., vol. 20, no. 11, pp. 1254–1259, 1998.

[8] Y. F. Ma, and H. J. Zhang, “Contrast-based image attention analysis by using fuzzy growing,” Eleventh ACM International Conference on Multimedia, Berkeley, Ca, Usa, November, pp:374-381, 2003.

[9] J. Harel, C. Koch, and P. Perona, “Graph-based visual saliency,” in NIPS, pp. 545–552, 2007.

[10] X. Hou and L. Zhang, “Saliency detection: A spectral residual approach,” in CVPR, pp. 1–8, 2007.

[11] R. Achanta, F. Estrada, P. Wils, and S. Susstrunk, “Salient region detection and segmentation,” in Comp. Vis. Sys., 2008.

[12] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, “Frequency- tuned salient region detection,” in CVPR, 2009.

[13] M. M. Cheng, G. X. Zhang, N. J. Mitra, X. Huang, and S. M. Hu, “Global contrast based salient region detection,” in CVPR, 2011.

[14] X. Shen and Y. Wu, “A unified approach to salient object detection via low rank matrix recovery,” in CVPR, 2012.

[15] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, “Saliency filters: Contrast based filtering for salient region detection,” in CVPR, 2012.

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

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

[18] Qin Y,Lu H,Xu Y,Wang H. “Saliency detection via cellular automata,” in CVPR, 2015.

[19] X. Zhang, C Xu, M Li, and R. K. F. Teng, “Study of visual saliency detection via nonlocal anisotropic diffusion equation,” Pattern Recognition, 48(4), 1315-1327, 2015.

[20] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Shum, “Learning to detect a salient object,” IEEE Trans. Patt. Anal. Mach. Intell., vol. 33, no. 2, pp. 353–367, 2011.

[21] Borji, A., Cheng, M. M., Jiang, H., & Li, J. (2014). “Salient object detection: a survey,”. Eprint Arxiv, 16(7), 3118.

[22] W. Einh ̈auser and P. K ̈onig. “Does luminance-contrast contribute to a saliency map for overt visual attention?” Eur JNeurosci, 17(5):1089–1097, Mar. 2003.

[23] D. Parkhurst, K. Law, and E. Niebur. “Modeling the role of salience in the allocation of overt visual attention.” Vision Res, 42(1):107–123, Jan. 2002.

[24] P. Reinagel and A. M. Zador. “Natural scene statistics at the centre of gaze.” In Network: Computation in Neural Systems, pages 341–350, 1999.

[25] R. Achanta, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, “SLIC Superpixels,” Tech. Rep. EPFL, 2010, Tech.Rep. 149300.