Object-based attention: saliency detection using contrast via background prototypes

Quan Zhou

An object-based attention model to predict visual saliency using contrast against the 'background prototypes' is presented. The proposed model automatically identifies a series of regions far away from the image centre as background prototypes. The visual saliency is then calculated using the colour contrast against these background prototypes. Promising experimental results demonstrate the effectiveness of the proposed model in terms of detection accuracy and implementation efficiency.

Introduction: The human attention system (HAS) is able to quickly detect the most interesting regions in a given scene. An important and open issue in computer vision is to simulate the HAS in human–machine interactions, which enables the hardware systems to automatically focus on and capture interesting target objects. Owing to the limited computational resources (less powerful processor and limited memory) of electronic equipments, the designed saliency models are expected to be easy to implement, while maintaining accurate and robust detection results. Therefore, considerable effort has been devoted to detecting salient regions over the past few years [1–7].

Existing saliency models could be categorised into two classes: top-down [8–10] and bottom-up [6, 11–13]. Top-down methods employ high-level cues (e.g. face); however, it is hardly generalised since their learning process often needs numerous computing resources. On the contrary, bottom-up approaches mainly estimate the foreground saliency based on simple low-level image features (e.g. luminance, colour and orientation), thus they are more convenient to apply to the practical scenarios of human–machine interactions. As a pioneer work, Itti *et al.* [11] introduced a biologically inspired saliency model based on the centre-surround operation. Graph-based saliency models [6, 12] are suggested to predict saliency following the principle of Markov random work theory. Some researchers attempted to detect irregularities such as visual saliency in the frequency domain [1, 13]. These saliency models, however, have limited capacity in complex scenes, where the salient and background regions are often heterogeneous.

Another instance of biological evidence shows that human visual attention is often attracted to the image centre [14], since the background tends to be located in the image boundary. Inspired by this biological fact, we propose a bottom-up saliency model relying on the contrast against 'background prototypes'. Here, 'background prototypes' mean the superpixels are located far away from the image centre, even in the boundary of an image. Another important aspect of saliency modelling is how to computationally measure visual saliency. The traditional models often estimate saliency using the centre-surround contrast [11] or the global contrast against the entire scene [10, 15]. On the contrary, we resort to calculating colour-based contrast with respect to the background prototypes. The experimental results show the effectiveness of our model in terms of more robust and accurate saliency maps and high computational efficiency.

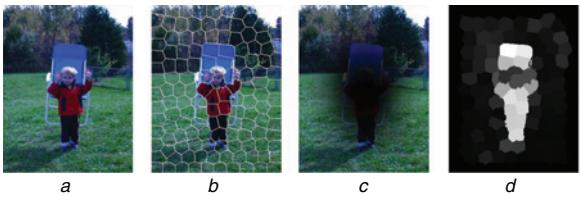

Fig. 1 *Overview of our method*

a Original image
b Image partition using over-segmentation algorithm [16]
c Image regions located in surrounding of image tend to be backgrounds
d Our saliency map

Proposed method: Fig. 1 illustrates an overview of our method. An input image (Fig. 1*a*) is first over-segmented into perceptually homogeneous superpixels (Fig. 1*b*). Considering the relative location with respect to the entire image, the regions, which locate far away from an image centre, are then selected as background prototypes (Fig. 1*c*). Finally, the visual saliency (Fig. 1*d*) is predicted by the colour contrast against the select background prototypes. Immediately below, we elaborate the details of the proposed model for saliency object detection.

The first step is to segment the input image into a set of superpixels using the method of [16]. There are two parameters to be tuned for this segmentation algorithm, namely (rgnSize, regulariser), which denote the number of superpixels used for over-segmentation and the trades-off appearance for spatial regularity, respectively. The advantage of using this technique is that it can often group the homogeneous regions with similar appearance and yet preserve the true boundaries of objects. We adopt a colour histogram as the feature to represent each superpixel, which is a robust global description of region appearance. The histogram is computed in the CIELab colour space because of its perceptual accuracy. As a result, the ith superpixel R_i is described by a K-dimensional colour histogram $\boldsymbol{h}_i$, where the kth element is denoted by $\boldsymbol{h}_i(k)$.

The second step is to select a series of robust background prototypes. According to [14], we generate a location prior map using the inverse Gaussian distribution from pixel $\boldsymbol{x} = (x, y)$ to image centre $\boldsymbol{c} = (c_x, c_y)$

$$p(\boldsymbol{x}) = 1 - \exp\{-d(\boldsymbol{x}, \boldsymbol{c})\} \tag{1}$$

where $d(\boldsymbol{x}, \boldsymbol{c}) = (x - c_x/2W)^2 + (y - c_y/2H)^2$ and W and H are the image width and height, respectively. To this end, the superpixel R_i is able to obtain a location score s_i as the mean value of $p(\boldsymbol{x})$, where $\boldsymbol{x}$ is contained in R_i. The superpixels, where the corresponding scores are higher than a predefined threshold θ, are selected as background prototypes (denoted by $R^B = \{R_1^B, R_2^B, \ldots, R_M^B\}$)

$$R_i^B = R_i, \quad \text{if} \quad s_i \geq \theta \tag{2}$$

with $\theta = 0.8$ empirically. Usually, the regions close to the image boundary are preferentially selected, which may be appropriate to represent the diversity of backgrounds.

Finally, the proposed saliency is evaluated by the global colour-based contrast. Instead of computing contrast with respect to the entire scene [15], we calculate the contrast in terms of the χ^2 distance from each superpixel to the selected background prototypes. The appearance contrast between R_i and background prototype $R_j^B \in R^B$ is defined as

$$C_{ij}(R_i, R_j^B) = \frac{1}{2}\sum_{k=1}^{K}\frac{(\boldsymbol{h}_i(k) - \boldsymbol{h}_j^B(k))^2}{\boldsymbol{h}_i(k) + \boldsymbol{h}_j^B(k)} \tag{3}$$

The salient regions are sometimes also heterogeneous (as shown in Figs. 1 and 3); nevertheless the colours of these regions are still quite distinct from those of the background prototypes. In addition, the background regions frequently have similar appearance as the background prototypes. Therefore, the saliency map of R_i is finally computed as the minimum value of contrast among all background prototypes

$$S(R_i) = \min_j C_{ij}(R_i, R_j^B) \tag{4}$$

where each image pixel belonging to R_i is assigned with a saliency value $S(R_i)$ and the final saliency map is normalised to a fixed range [0, 1].

Experimental results: To demonstrate the effectiveness of our model, experiments were conducted on the publicly available dataset [13, 15], which is widely used for evaluation. Six saliency models are selected as baselines for comparison, which are visual attention measure (IT [11]), graph-based saliency (GB [12]), frequency-tuned saliency (FT [13]), context-aware saliency (CA [8]), global-contrast saliency (RC [15]) and midlevel saliency (MS [6]). In our implementation, the segmentation parameters (rgnSize, regulariser) are chosen to be (30, 1.0) and each colour histogram is a $\boldsymbol{K} = 60$ dimensional vector for each superpixel. For the methods of IT, GB, FT and RC, we implement the authors' code using a Dual Core 2.6 GHz machine with 4 GB memory to generate saliency maps, whereas the saliency maps of other methods are provided by [6, 15]. Table 1 reports the comparison results of average running time between the proposed method and other baselines, and demonstrates that the proposed approach achieves high computational efficiency.

For quantitative evaluations, all the models are evaluated by the widely used criteria [13, 15], including precision, recall and

F-measure. In the first evaluation, fixed thresholds varying from 0 to 255 are used to segment the normalised saliency maps. A series of precision–recall pairs are obtained and the results are reported in terms of precision–recall curve, as shown in Fig. 2*a*. It confirms that our method outperforms other models. In the second evaluation, we perform binary segmentation on the saliency maps using adaptive thresholding [13]. The precision and recall values are then calculated and the F-measure is also obtained for evaluation with

$$F = ((1 + \beta^2) \times P \times R)/(\beta^2 \times P + R) \quad (5)$$

where P is the precision, R is the recall and $\beta^2 = 0.3$ following [13, 15]. The results are comparatively plotted in Fig. 2*b*. Our method achieves the best precision, recall and F-measure. The improvement of recall over other methods is more significant, which indicates that our model is likely to detect more salient regions with a high accuracy.

Table 1: Comparison of average running time (most images are with resolution 400 × 300)

Method	IT [11]	GB [12]	FT [13]	CA [8]	RC [15]	MS [6]	Ours
Time (s)	0.611	1.614	0.016	1.483	0.253	0.77	0.174
Code	MATLAB	MATLAB	C++	C++	C++	MATLAB	MATLAB

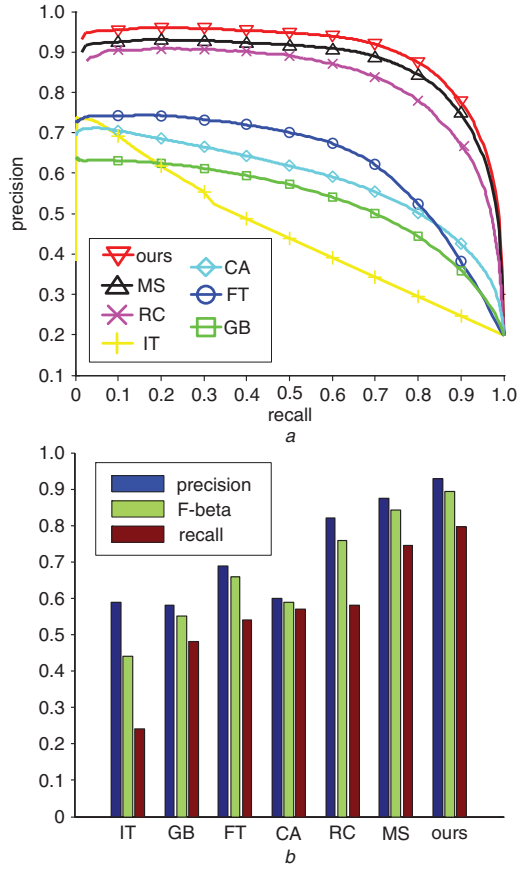

Fig. 2 *Quantitative comparison on MSRA 1000 dataset*

a Precision–recall curves of our method compared with CA [8], IT [11], GB [12], FT [13], RC [15] and MS [6]
b Average precision, recall and F-measure with adaptive-thresholding segmentation

For qualitative comparison, some visual saliency maps obtained by the various methods are shown in Fig. 3. Our model is able to yield high-resolution saliency maps, which are closest to the ground truth. By contrast, the other models produce overblurred saliency maps suffering from loss of object shape information (e.g. IT and GB), over-emphasising object boundaries (e.g. CA) or highlighting the cluttered background regions (e.g. FT, RC and MS).

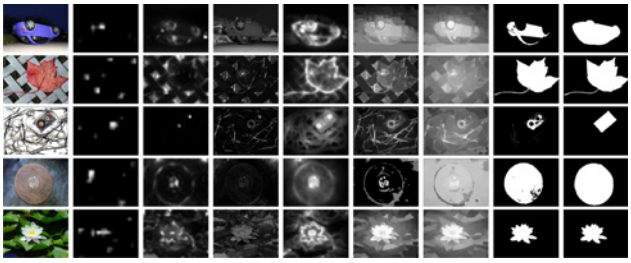

Fig. 3 *Some examples of visual comparison of previous approaches with our method*

From left to right: original image, results from IT [11], GB [12], FT [13], CA [8], RC [15], MS [6], proposed method and ground truth

Conclusion: A novel approach is proposed to perform salient object detection with attractive performance. The proposed approach utilises the colour-based contrast measurement with respect to background prototypes to predict visual saliency. The proposed method is able to produce a saliency map that is more consistent with the HAS than a number of existing approaches [6, 8, 11–13, 15], as demonstrated in our experiments. There are several issues that need to be further exploited. The first issue is how to predict the salient object when they locate in an image boundary, since only centre-biased consumption is considered in this Letter. Secondly, we are interested in extending our model to estimate region saliency in the spatio-temporal domain (e.g. video sequence).

Acknowledgments: This work was supported by the National Natural Science Foundation of China (grant nos 61201165, 61271240) and the Scientific Research Foundation of Nanjing University of Posts and Telecommunications (grant no NY213067).

8 May 2014
doi: 10.1049/el.2014.0903
One or more of the Figures in this Letter are available in colour online.

Quan Zhou (*College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, 66 Xin Mofan Road, Nanjing 210003, People's Republic of China*)

E-mail: quan.zhou@njupt.edu.cn

References

1 Wu, D., Sun, X.D., Jiang, Y.Y., and Hou, C.F.: 'Unsupervised visual saliency detection via information content measuring', *Electron. Lett.*, 2012, **48**, (25), pp. 1591–1593
2 Ko, B., and Nam, J.: 'Object-of-interest image segmentation based on human attention and semantic region clustering', *J. Opt. Soc. Am. A*, 2006, **23**, (10), pp. 2462–2470
3 Xie, Y., Lu, H., and Yang, M.: 'Bayesian saliency via low and mid level cues', *IEEE Trans. Image Process*, 2013, **22**, (5), pp. 1689–1698
4 Ma, L., Tian, J., and Yu, W.: 'Visual saliency detection in image using ant colony optimisation and local phase coherence', *Electron. Lett.*, 2010, **46**, (15), pp. 1066–1068
5 Li, J., Levine, M.D., An, X., Xu, X., and He, H.: 'Visual saliency based on scale-space analysis in the frequency domain', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2013, **35**, (4), pp. 996–1010
6 Yu, J., and Tian, J.: 'Saliency detection using midlevel visual cues', *Opt. Lett.*, 2012, **37**, (23), pp. 4994–4996
7 Wang, W., Wang, Y., Huang, Q., and Gao, W.: 'Measuring visual saliency by site entropy rate'. Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2010, pp. 2368–2375
8 Goferman, S., Zelnik-Manor, L., and Tal, A.: 'Context-aware saliency detection', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2012, **34**, (10), pp. 1915–1926
9 Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., and Shum, H-Y.: 'Learning to detect a salient object', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011, **33**, (2), pp. 353–367
10 Torralba, A.: 'Modeling global scene factors in attention', *J. Opt. Soc. Am. A*, 2003, **20**, (7), pp. 1407–1418
11 Itti, L., Koch, C., and Niebur, E.: 'A model of saliency-based visual attention for rapid scene analysis', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1998, **20**, (11), pp. 1254–1259
12 Harel, J., Koch, C., and Perona, P.: 'Graph-based visual saliency'. Proc. of Advances in Neural Information Processing Systems, 2006, pp. 545–552

13 Achanta, R., Hemami, S., Estrada, F., and Süsstrunk, S.: 'Frequency-tuned salient region detection'. Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2009, pp. 1597–1604
14 Tatler, B.: 'The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions', *J. Vis.*, 2007, **7**, (14), pp. 1–17
15 Cheng, M., Zhang, G., Mitra, N., Huang, X., and Hu, S.: 'Global contrast based salient region detection'. Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2011, pp. 409–416
16 Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Susstrunk, S.: 'SLIC superpixels compared to state-of-the-art superpixel methods', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2012, **34**, (11), pp. 2274–2282