2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)
Banff Center, Banff, Canada, October 5-8, 2017

Focusness Guided Salient Object Detection

Xiaolin Xiao, Yicong Zhou
Department of Computer and Information Science
University of Macau, Macau 999078, China
Email: shellyxiaolin@gmail.com, yicongzhou@umac.mo

***Abstract*—Salient object detection aims to correctly highlight the most salient object(s) in an image. Combining fine-grained contrast prior with rough-grained object consistency, this paper proposes a Focusness Guided Salient object detection (FGS) algorithm. To obtain clean and precise contrast map, FGS uses the focusness prior to guide the contrast map. Combing different saliency priors, FGS utilizes a unified least-square framework to generate the final optimal salient map. Experiments demonstrate the proposed method outperforms the state-of-the-arts.**

I. Introduction

Origins from visual attention, saliency detection can be grouped into two different categories [1], [2]. The first one is eye fixation prediction, emphasizing on highlighting the points that human eyes prefer at first glance. The other is salient object detection with the purpose of detecting the most salient entire object(s) in an image. Compared with eye fixation prediction, the task of salient object detection involves some higher level attention beyond first glance. This is because our eyes need complex cognition to identify the entire structure and meaning of an "object", instead of just some perceptual sensitive points. This paper aims to correctly popping up the complete salient object(s). Recently, salient object detection greatly prompts the computer vision applications and also the artificial intelligence tasks. Examples can be found in scene parsing [2], automatic image cropping [3], object tracking [4], object importance evaluation [5], image and video interestingness [6], and human-mechine interaction [7].

The processing units for salient object detection can be pixels, patches, superpixels, or regions. Salient object detection is widely considered as capturing the uniqueness, or the rarity of an image. And uniqueness is widely calculated as the pixel-wise center-surround contrast [8]. To reduce the computation cost from pixel-wise comparison, block/patche-based methods are proposed [9]. However, the grid block/patch representation lacks of adherence with actual object boundaries. Thus, it severely reduces the performance of salient object detection algorithms. To overcome this problem, superpixel (compact superpixel) [10], [11] and/or rough region (uncompact superpixel) [12], [13] based methods emerge. Both of superpixels and regions have advantage in capturing the object boundaries, while reducing the computational cost by using smaller processing units. And they vary in the following aspects: (1) superpixel representation [14] advances in capturing bottom-up saliency since they can evaluate the uniqueness in small scale; (2) region representation [15], [16] benefits in identifying the whole object(s), instead of just part of it(them).

Ever since the emergence of saliency detection, many empirical priors and statistical models are exploited. The empirical priors comes from psychology or neurobiology, engaging to exploit the characteristics of human visual systems. Common empirical priors have been widely used, and their effectiveness is demonstrated. Among them, center-surround contrast [12], [17], [18] is the most widely used cue. It insists that units that have high contrast with their surroundings are salient. Besides, background prior [19]–[21] is involved since units along image borders are less likely to be salient. Foucsness [13], [22], [23] is also a valuable cue which assumes that salient object should be photographed in focus. Meantime, statistical models make use of mathematical techniques to analyze natural images, and to guide saliency detection. Examples are the gaussian mixture model [14], the low rank model [20], the Bayesian inference model [21], and many others.

Together with the previous bottom-up saliency cues, higher level priors are also exploited. The most popular top-down saliency prior is object prior. Objects vary rapidly in different categories, thus, their properties are usually measured in supervised way, either category-dependent object detection mechanism [24], or category-independent object classifiers [22] are trained. The drawback for supervised object prior lies in the cost from training.

Based on the above observations, we propose a Focusness Guided Salient object detection (FGS) algorithm. Our contributions are listed as follows:

- Combining fine-grained superpixels and rough regions, FGS is able to mutually achieve precise salient priors.
- Using the focusness cue to guide the contrast map, FGS can obtain object level consistency.
- Integrating the bottom-up salient priors and focusness map, FGS utilizes a unified optimization framework to generate the final optimal salient map.

II. Focusness Guided Saliency (FGS)

A. Algorithm Structure

The framework of FGS is illustrated in Fig. 1. Given an input image I, first, it is segmented into two scales, fine-grained superpixels and rough regions. The superpixels work as atomic units to identify bottom-up contrast and background probability, while the rough regions are used to detect the focusness of objects in image. Based on the assumption that the most salient objects usually capture the focused part on an image, we further use the focusness map to guide the contrast

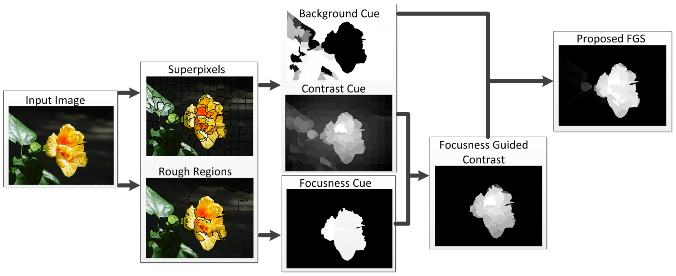

Fig. 1: Flow chart of the proposed FGS algorithm.

map, overwhelming high contrast in unfocused regions. Thus, it results in a more precise and cleaner foreground prior, that is, the focusness guided contrast prior. Finally, the salient object map is optimized combing the focusness guided contrast map and the background probability map.

To formulate our problem, the following symbols are adopted, $\{s_1, s_2, ..., s_M\}$ represents the superpixel segments, $\{r_1, r_2, ..., r_N\}$ represents the rough regions.

B. Background Cue

The state-of-the-art salient object detection methods exploit boundary cue to exclude the non-salient part of the image, and thus, to avoid the false positive detection. The boundary cue assumes that superpixels/regions lie on image borders are less likely to be salient. Following the idea in [10], we exploit the robust background connectivity prior on fine-grained superpixels $\{s_1, s_2, ..., s_M\}$ to measure the background probability. For each superpixel s_i, the boundary connectivity is defined as:

$$BonCon(s_i) = \frac{L(s_i)}{A(s_i)}, \tag{1}$$

where $L(\cdot)$ and $A(\cdot)$ represent the length and spanning area of a superpixel along the image border, respectively. Thus, the background probability is measured using:

$$w_{bg}(s_i) = 1 - exp(-\frac{BonCon^2(s_i)}{2\sigma^2_{BonCon}}), \tag{2}$$

where σ^2_{BonCon} is the parameter that controls the influence of $BonCon$ on w_{bg}. The background probability w_{bg} is used as our background cue. It is large when $BonCon$ is large, and vice versa.

C. Focusness Cue

Focusness cue claims that a salient object is often photographed in focus to attract human attention. Usually, focusness is measured from the degree of focal blur along the edges of objects. Then, edge blur is spread to the whole image. We exploit a fast and effective way to estimate the focusness of images on a rough segmentation. The rough regions $\{r_1, r_2, ..., r_N\}$ instead of the fine-grained superpixels are used since the concept of focusness is highly connected to objects, whereas, the fine-grained superpixels can hardly provide object level consistency. The focusness $\{f(r_p)\}_{p=1}^N$ can be measured in three steps: (1) pixel level edge strength detection, then (2) focusness estimation on detected edges, and finally, (3) region level focusness spread.

The focusness cue provides a rough estimation on the focusness of objects in a image, however, it is prone to generate high false positive estimation since we use very large regions, greatly weakening the performance. To avoid this problem, we make use of the background probability cue to refine the focusness map. Our principle is that, background regions are less likely to be focused, whereas, foreground regions are preferred to be in focus. For $\{r_1, r_2, ..., r_N\}$, we calculate its background connectivity level $\{bg(r_p)\}_{p=1}^N$ using the average background probability from its composed pixels, while the background probability of each pixel is set as its corresponding superpixels. And then, we discard the regions that are highly connected to image borders.

$$bg(r_p) = \begin{cases} bg(r_p), & bg(r_p) < T \\ 0, & bg_{r_p} = 0 \end{cases} \tag{3}$$

where $T = \max\{\frac{\sum bg(r_p)}{2N}, 0.2\}$ is the threshold. Note that the threshold T is set to be smaller than a fixed value 0.2 to avoid much false negtive prediction. Finally, the region level refined focusness $\{rf(r_p)\}_{p=1}^N$ is achieved as:

$$rf(r_p) = bg(r_p) * f(r_p). \tag{4}$$

D. Contrast Cue

The contrast cue, also called the center-surround cue, is the most widely used approach to identify saliency [12], [17], [18]. It is inspired from neuroscience that human eyes prefer the "different" part of an image. Specifically, if a unit is different (regarding to different features, e.g., color, texture) from its surroundings, it should be salient. Such a unit can be implemented by a pixel, a block, or a superpixel, a region. Our contrast prior is constructed on fine-grained superpixels $\{s_1, s_2, ..., s_M\}$, according to color difference and spatial distance:

$$Ctr(s_i) = \sum_{j=1}^{M} w_s(s_i, s_j) d_c(s_i, s_j), \tag{5}$$

where $w_s(s_i, s_j)$ is the spatial weight from superpixel s_j to s_i, and $d_c(s_j, s_j)$ is the color distance between s_i and s_j. In our experiments, color distance is calculated using the Lab distance using CIE76 formula, and $w_s(s_i, s_j)$ is set as:

$$w_s(s_i, s_j) = exp(-\frac{d_s(s_i, s_j)}{2\sigma_s^2}), \tag{6}$$

where σ_s^2 is the spatial parameter. In our experiments, σ_s is set to 0.4 so that we can mainly focus on the surrounding superpixels within two or three layers.

E. Focusness Guided Contrast

For $\{s_1, s_2, ..., s_M\}$, we calculate its focusness strength $\{rf_{s_i}\}_{i=1}^M$ with the average focusness values of its composed

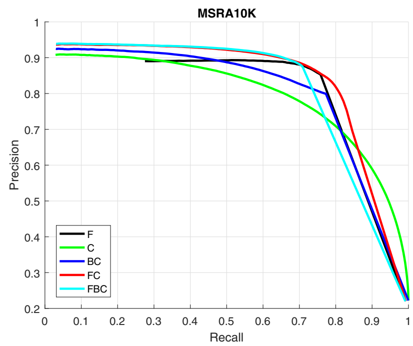

Fig. 3: The PR curve of different foreground priors.

TABLE I: MAE of different saliency detection algorithms.

	FGS	GMR	HC	RBD	RC	SF	UFO
MSRA10K	0.1107	0.1257	0.2156	**0.1083**	0.1372	0.1753	0.1496
SED2	**0.1118**	0.1630	0.1932	0.1209	0.1478	0.1794	0.1803
PASCAL-S	**0.1935**	0.2330	0.3536	0.2013	0.3015	0.2533	0.2464
ECSSD	**0.2206**	0.2366	0.2562	0.2274	0.2351	0.2737	0.2558

TABLE II: F_β of different saliency detection algorithms.

	FGS	GMR	HC	RBD	RC	SF	UFO
MSRA10K	**0.761**	0.7657	0.5982	0.7607	0.6914	0.5565	0.6704
SED2	**0.7595**	0.7111	0.6477	0.7413	0.718	0.5835	0.636
PASCAL-S	**0.5597**	0.5589	0.3715	0.5542	0.3011	0.3812	0.4651
ECSSD	0.5928	**0.6124**	0.4165	0.5903	0.5823	0.4031	0.5106

pixels, and use them to guide (filter) the contrast. The guided contrast is:

$$w_{fg}(s_i) = rfCtr(s_i) = rf(s_i) * Ctr(s_i). \tag{7}$$

F. Optimization

Our objective function is constructed by minimizing the energy of three components: namely, foreground term, background term, and a smooth term:

$$E = \sum_{i=1}^{M} w_{fg}^2(s_i) + \sum_{i=1}^{M} w_{bg}^2(s_i) + \sum_{i=1}^{M}\sum_{j=1}^{M} w_{sm}^2(s_i, s_j), \tag{8}$$

where the smooth term $w_{sm}(s_i, s_j)$ is set as the Lab color distance among each adjacent superpixel pair, and the adjacent is calculated on a two layer neighborhood, together with the all border superpixels as a connected layer. Note that $w_{sm}(s_i, s_j)$ is larger when s_i and s_j share similar features, indicating that similar superpixels should have close salient values; and $w_{sm}(s_i, s_j)$ is close to 0 when s_i and s_j have distinct color features. We enforce such a smooth only in a two layer neighboring regions, this is different from the contrast cue, which is measured on a larger (global) region. Minimizing energy E using least-square optimization, we can obtain the final salient map $\{\xi(s_i)\}_{i=1}^{M}$.

III. Experiments

To test the effectiveness of the proposed FGS, we use four standard datasets, MSRA10K [25], SED2 [26], PASCAL-S [27], and ECSSD [28]. Comparison metrics are the commonly used Precision-Recall (PR) curve, Receiver Operating Characteristics (ROC) curve, F_β measure, and Mean Absolute Error (MAE) [29].

A. Evaluation of Focusness Guided Contrast

First, we evaluate the performance of the proposed focusness guided contrast cue (FC). The PR curve of different foreground cues are plotted in Fig. 3. We show the effectiveness of Focusness (F), Contrast (C), and three ways to filter the contrast, namely, Background weighted Contrast (BC), Foceueness guided Contrast (FC), and Background and Focusness weighted Contrast (BFC). The result shows that FC outperforms the other foreground cues. We also provide a visual comparison to show the performance of different cues and their influence on the final salient maps in Fig. 2.

B. Comparison with the State-of-the-arts

Finally, we compare FGS with the state-of-the-arts. The competing algorithms are: GMR [19], HC [12], RBD [10], RC [12], SF [11], and UFO [13]. Fig. 4 illustrated the statistical comparison of different algorithms on four different datasets. For the PR curve, FGS outperforms the other methods on the MSRA10K and SED2 datasets, while UFO has the best PR curve on PASCAL-S, and RMR and RC have relative better PR curves on ECSSD. As to the ROC curve, FGS has the best results on all datasets. TABLE I and TABLE II show the MAE and F_β comparisons. Finally, the visualization of different algorithms are presented in Fig. 5. As to the computational cost, the proposed method can process an image of size $400*300$ within 3s on a Core 3.40 GHz machine with 8 GB RAM.

IV. Conclusion

In this paper, we proposed a Focusness Guided Salient object detection (FGS) algorithm. FGS exploits two different processing scales, working mutually to generate clean and precise salient priors. The region-based focusness prior is used to guide the contrast map, improving the detection performance by implicitly incorporating object level consistency. Finally, FGS integrates the contrast, background, and focusness priors using a unified optimization framework to attain the final optimal salient map. Experiments demonstrated the advantages of FGS over the state-of-the-arts.

Acknowledgment

This work was supported in part by the Macau Science and Technology Development Fund under Grant FDCT/016/2015/A1 and by the Research Committee at University of Macau under Grants MYRG2014-00003-FST and MYRG2016-00123-FST.

References

[1] Neil DB Bruce and John K Tsotsos, "Saliency, attention, and visual search: An information theoretic approach," *Journal of vision*, vol. 9, no. 3, pp. 5–5, 2009.

Fig. 2: Visualization of different components and their effect on the final salient maps.

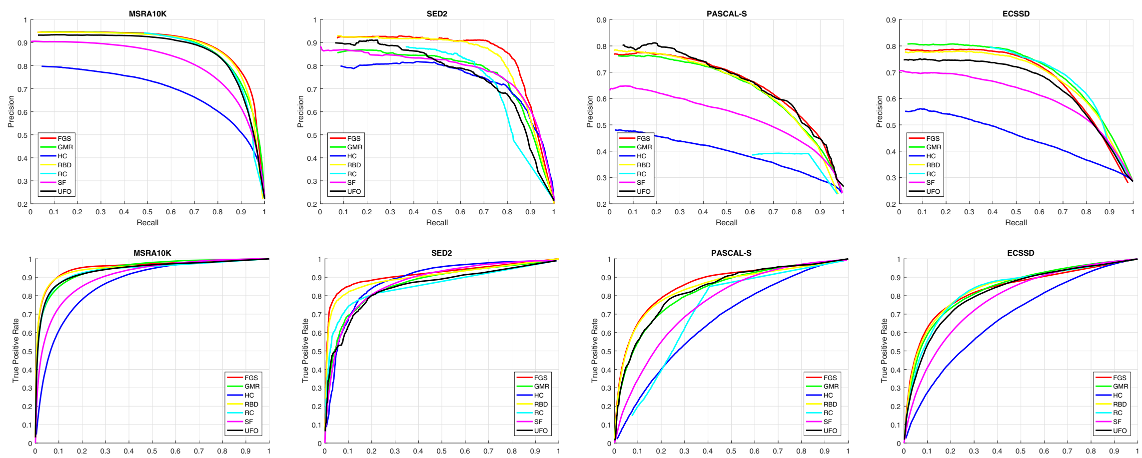

Fig. 4: PR curve and ROC curve of different saliency detection algorithms on four datasets. First row, PR curves; second row, ROC curves.

Fig. 5: Visualization of different saliency detection algorithms on different datasets.

[2] Ali Borji, Ming-Ming Cheng, Huaizu Jiang, and Jia Li, "Salient object detection: A survey," *arXiv preprint arXiv:1411.5878*, 2014.
[3] Anthony Santella, Maneesh Agrawala, Doug DeCarlo, David Salesin, and Michael Cohen, "Gaze-based interaction for semi-automatic photo cropping," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 2006, pp. 771–780.
[4] Xi Li, Weiming Hu, Chunhua Shen, Zhongfei Zhang, Anthony Dick, and Anton Van Den Hengel, "A survey of appearance models in visual object tracking," *ACM transactions on Intelligent Systems and Technology (TIST)*, vol. 4, no. 4, pp. 58, 2013.
[5] Merrielle Spain and Pietro Perona, "Measuring and predicting object importance," *International Journal of Computer Vision*, vol. 91, no. 1, pp. 59–76, 2011.
[6] Harish Katti, Kwok Yang Bin, Tat Seng Chua, and Mohan Kankanhalli, "Pre-attentive discrimination of interestingness in images," in *2008 IEEE International Conference on Multimedia and Expo*. IEEE, 2008, pp. 1433–1436.
[7] Yusuke Sugano, Yasuyuki Matsushita, and Yoichi Sato, "Calibration-free gaze sensing using saliency maps," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 2667–2674.
[8] Feng Liu and Michael Gleicher, "Region enhanced scale-invariant saliency detection," in *Multimedia and Expo, 2006 IEEE International Conference on*. IEEE, 2006, pp. 1477–1480.
[9] Dominik A Klein and Simone Frintrop, "Center-surround divergence of feature statistics for salient object detection," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2214–2219.
[10] Wangjiang Zhu, Shuang Liang, Yichen Wei, and Jian Sun, "Saliency optimization from robust background detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 2814–2821.
[11] Federico Perazzi, Philipp Krähenbühl, Yael Pritch, and Alexander Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 733–740.
[12] Ming-Ming Cheng, Niloy J Mitra, Xiaolei Huang, Philip HS Torr, and Shi-Min Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2015.
[13] Peng Jiang, Haibin Ling, Jingyi Yu, and Jingliang Peng, "Salient region detection by ufo: Uniqueness, focusness and objectness," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1976–1983.
[14] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.
[15] Dorin Comaniciu and Peter Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 5, pp. 603–619, 2002.
[16] Andrea Vedaldi and Stefano Soatto, "Quick shift and kernel methods for mode seeking," in *European Conference on Computer Vision*. Springer, 2008, pp. 705–718.
[17] Qiong Yan, Li Xu, Jianping Shi, and Jiaya Jia, "Hierarchical saliency detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1155–1162.
[18] Christian Scharfenberger, Alexander Wong, Khalil Fergani, John S Zelek, and David A Clausi, "Statistical textural distinctiveness for salient region detection in natural images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 979–986.
[19] Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang, "Saliency detection via graph-based manifold ranking," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3166–3173.
[20] Xiaohui Li, Huchuan Lu, Lihe Zhang, Xiang Ruan, and Ming-Hsuan Yang, "Saliency detection via dense and sparse reconstruction," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2976–2983.
[21] Bowen Jiang, Lihe Zhang, Huchuan Lu, Chuan Yang, and Ming-Hsuan Yang, "Saliency detection via absorbing markov chain," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1665–1672.
[22] Yangqing Jia and Mei Han, "Category-independent object-level saliency detection," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 1761–1768.
[23] Kai-Yueh Chang, Tyng-Luh Liu, Hwann-Tzong Chen, and Shang-Hong Lai, "Fusing generic objectness and visual saliency for salient object detection," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 914–921.
[24] Xiaohui Shen and Ying Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 853–860.
[25] Vida Movahedi and James H Elder, "Design and perceptual validation of performance measures for salient object segmentation," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010, pp. 49–56.
[26] Tie Liu, Zejian Yuan, Jian Sun, Jingdong Wang, Nanning Zheng, Xiaoou Tang, and Heung-Yeung Shum, "Learning to detect a salient object," *IEEE Transactions on Pattern analysis and machine intelligence*, vol. 33, no. 2, pp. 353–367, 2011.
[27] David R Martin, Charless C Fowlkes, and Jitendra Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 5, pp. 530–549, 2004.
[28] Huaizu Jiang, Jingdong Wang, Zejian Yuan, Yang Wu, Nanning Zheng, and Shipeng Li, "Salient object detection: A discriminative regional feature integration approach," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 2083–2090.
[29] Ali Borji, Ming-Ming Cheng, Huaizu Jiang, and Jia Li, "Salient object detection: A benchmark," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5706–5722, 2015.