SALIENCY DETECTION USING A BACKGROUND PROBABILITY MODEL

Junling Li[1], *Fang Meng*[1*], *Yichun Zhang*[2]

[1]Communication University of China, Beijing, China
[2]China Art Science and Technology Institute, Beijing, China

ABSTRACT

Image saliency detection has been long studied, while several challenging problems are still unsolved, such as detecting saliency inaccurately in complex scenes or suppressing salient objects in the image borders. In this paper, we propose a new saliency detection algorithm in order to solving these problems. We represent the image as a graph with superpixels as nodes. By considering appearance similarity between the boundary and the background, the proposed method chooses non-saliency boundary nodes as background priors to construct the background probability model. The probability that each node belongs to the model is computed, which measures its similarity with backgrounds. Thus we can calculate saliency by the transformed probability as a metric. We compare our algorithm with ten-state-of-the-art salient detection methods on the public database. Experimental results show that our simple and effective approach can attack those challenging problems that had been baffling in image saliency detection.

***Index Terms*—**visual saliency, background probability, boundary knowledge, background priors.

1. INTRODUCTION

Saliency detection is automatically to make the most interesting objects or regions pop out with respect to its neighbors in the entire image. It benefits numerous applications in computer vision, including object recognition [1], content-aware image retargeting [2], image segmentation [3], adaptive compression [4].

In general, all bottom-up saliency methods depend on some prior knowledge of the entire image, which can be characterized from different perspectives, such as contrast, rarity, repeated, etc. Itti et al. [5] proposed a highly influential computational method using local center-surrounded contrasts. A lot of existing algorithms are developed from this algorithm, such as MA and Zhang et al. [6], Achanta et al. [7], Harel et al. [8], and Frintrop et al. [9]. Fourier spectrum analysis in the frequency domain has been used to detect visual saliency [10]-[11]. Achanta et al. [12] introduced a global saliency to the average image color. This approach uses color distance to detect saliency map from any given pixel color salient objects or complex backgrounds (see Fig.1 (b)). Recently, saliency detection methods based on the background model have been proposed. Jiang et al. [13] formulated a saliency detection algorithm based on backgrounds by using the time property in an absorbing Markov chain, where all boundary nodes were directly set as absorbing nodes. While this method outperforms most existing methods in terms of precision and recall, it may fail to correctly highlight the salient regions when salient objects locate in the image boundaries. According to the computation of the absorbed time, most parts of the salient object touching image boundaries are suppressed (see Fig.1 (c)). Reference [13] provides details of this method and some typical failure cases.

(a) (b) (c) (d)

Fig. 1 Examples of saliency detection results. (a) Input images, (b)-(d) Results of Achanta et al. [12], Jiang et al. [13], Ours.

In this paper, we propose a simple and efficient saliency detection approach based on a background probability model. Different from the existing methods based on backgrounds, the main contribution of our work is to adopt FT method [12] to compute background prior knowledge instead of relying on the hypothesis that regards all boundary knowledge as background knowledge directly. This approach is to contribute to detect salient objects in the image borders. As shown in Fig.1, our method can effectively distinguish salient regions and backgrounds, even in the presence of complex backgrounds or salient objects in the image borders.

* Corresponding author's email: mengfang@cuc.edu.cn. This work is partially supported by the University Level Academic Program Development of Communication University of China.

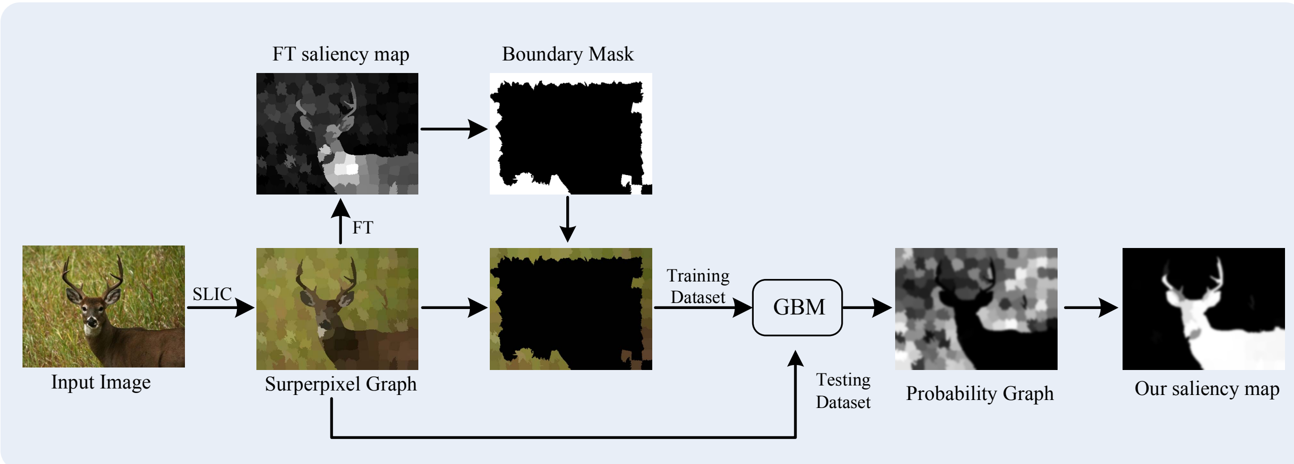


Fig.2 Diagram of our proposed model

2. THE PROPOSED SALIENCY EXTRACTION MODEL

In order to capture perceptually homogeneous regions, we represent the input image as a graph with 200 superpixel nodes generated by the SLIC algorithm [14]. Each node has the average feature of all pixels in the corresponding superpixel node. Thus the graph retains the important structure information and color information while reducing unnecessary image details. Inspired by the drawbacks of [13], we firstly determine whether boundary prior belongs to background prior. Then, we select the boundary nodes judged as background to construct the Gaussian Background Model (GBM). The probability distribution that each node belongs to GBM can be future computed. Lastly, we measure the saliency of nodes using the probability distribution. The flowchart shown in Fig.2 demonstrates the framework of our saliency detection method in detail.

2.1. Selecting boundary nodes as background knowledge

Generally, the background regions often have similar appearance connectivity with image boundaries. Hence, many saliency algorithms based on backgrounds directly treat boundary knowledge as background knowledge. However, most parts of salient objects will be suppressed as background when they are located in the image boundaries. Our algorithm avoids this drawback by determining whether boundary nodes belong to background in advance and then choosing appropriate boundary nodes as background.

Owing to the high speed and accuracy of FT method [12], we exploit it to detect salient regional location roughly. FT method uses color distance to detect saliency values from any given node color to the average color of nodes in the CIELab space. The indexes of boundary nodes can be easily retrieved through scanning the full resolution label image, which can be obtained from the SLIC processing. Then, we can obtain the binary value of boundary nodes BN_ind , which can be denoted as:

$$BN\_ind(i) = \begin{cases} 0 & S_{FT}(i) \geq T \\ 1 & S_{FT}(i) < T \end{cases} \qquad i = 1,2,\ldots,t, \quad (1)$$

where i indexes the node around image borders, t is the number of all the boundary nodes, T represents twice the average of S_{FT} and S_{FT} denotes the normalized saliency value computed by FT method.

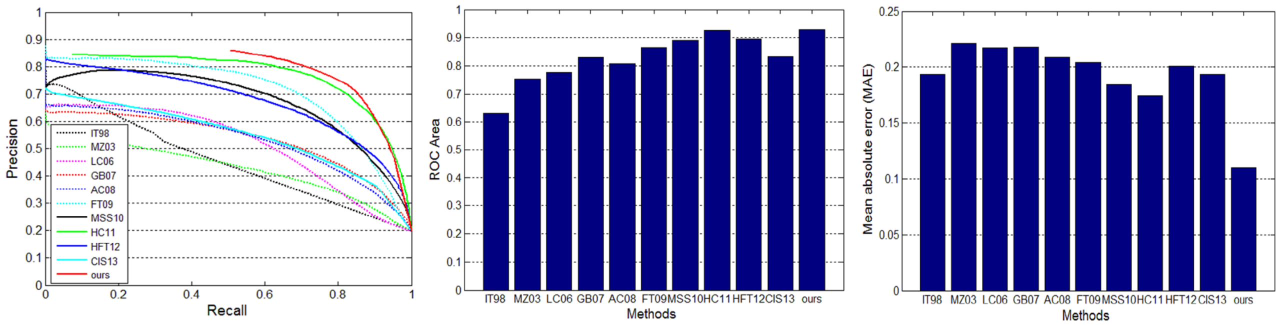
We define the nodes corresponding to non-zero elements of BN_ind as non-saliency boundary nodes. Using the Lab color space, we can extract the feature vector $[L,a,b]^T$ of non-saliency boundary nodes. In particular, it is necessary for us to note that the saliency result computed by FT method is incorrectly when the number of non-zero elements of BN_ind is less than half of t. In this case, we should inverse the values of BN_ind and then reselect the appropriate boundary nodes.

Fig.3 shows the effectiveness of selecting boundary nodes. As shown in Fig.3, we can conclude that the selection strategy of boundary nodes can contribute to the effect of the final saliency result.

(a) (b) (c)

Fig.3 Saliency maps using different boundary nodes. (a) Input image, (b) result of using all the boundary nodes, (c) result of using non-saliency boundary nodes.

2.2. Constructing the Gaussian Background Model



(a) (b) (c)

Fig.4 Objective comparison with ten alternative saliency methods using the 1000images: (a) the average Precision-Recall curves, (b) the average ROC Area, (c) the average MAE.

We use a training dataset $Q=\{q_1,q_2,\ldots,q_i,q_{i+1},\ldots,q_n\}$ to represent the selected boundary nodes, where n is the total number and q_i denotes a feature vector $[L_i,a_i,b_i]^T$. The dataset Q is applied to train a Gaussian Mixture Model (GMM) as GBM by Expectation-Maximization. After that, we can obtain the means, covariance matrixes and weights of GBM. Similarly, we define a dataset $P=\{p/p\in \textit{all nodes in the graph}\}$ and p denotes a feature vector $[L,a,b]^T$. Then, we compute the probability distribution that each node in the graph belongs to the GBM by (2) and (3).

$$P(p)=\sum_{k=1}^{K}\pi_k N(p;\mu_k;\Sigma_k) \tag{2}$$

$$N(p;\mu_k;\Sigma_k)=\frac{1}{\sqrt{2\pi|\Sigma_k|}}\exp\left[-\frac{1}{2}(p-\mu_k)^T\Sigma^{-1}(p-\mu_k)\right] \tag{3}$$

where μ_k , Σ_k , π_k represent the mean, covariance matrix, the weight of the k^{th} component, and K is the number of components. However, experiments found that our method has often a better effect when K is 1. Therefore, we adopt the Gaussian Single Model as GBM in this work.

As showed in Fig. 2, the highlighted points in the probability graph correspond to the points of large probability value, that is to say, these regions have more similar appearance with background regions.

2.3. Computing saliency

Therefore, our method of calculating the saliency map S for a graph can be formulated as:

$$S(i)=\frac{1}{2\sigma^2}\exp\left(-\frac{P(i)}{2\sigma^2}\right) \qquad i=1,2,\ldots,N \tag{4}$$

where i indexes the node in the graph, N is the number of nodes in the graph, and P denotes the probability of belonging to the GBM. We empirically set $\sigma=12$ in this work.

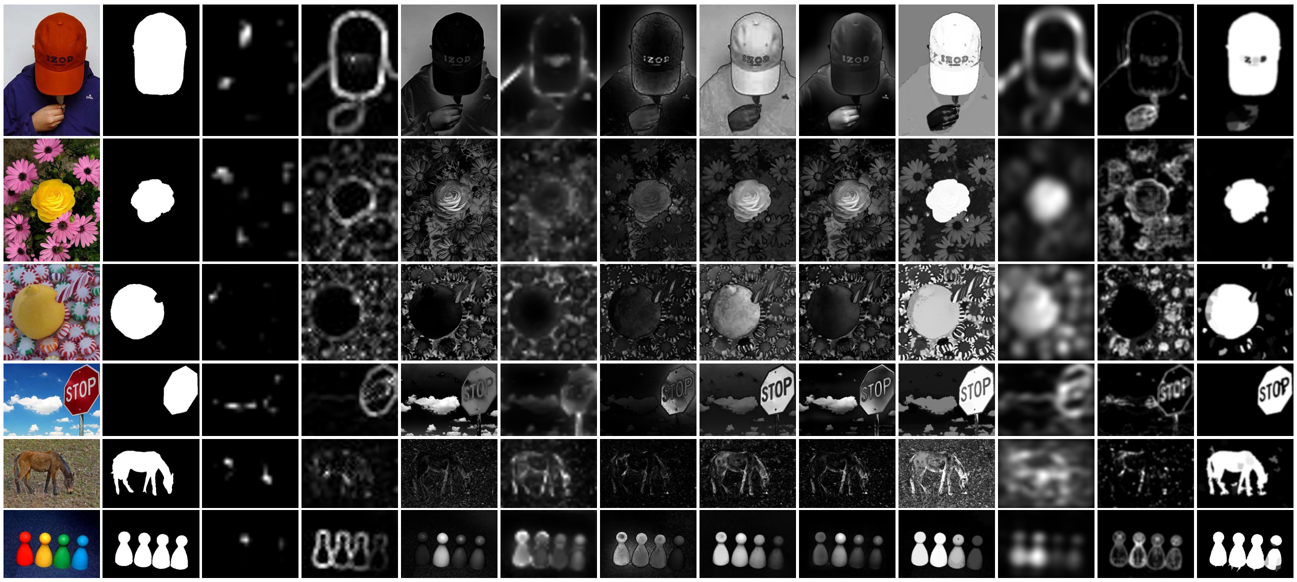
Finally, we normalize the saliency map S and output the saliency map with full resolution. From Fig.3(c), we can see the salient object is detected correctly and highlighted uniformly. Note that our method spends most of the execution time on SLIC processing, while the actual execution time of detecting saliency is only about 0.019 seconds per image which has resolution 400*300.

3. Experiments

We evaluate our method on 1000 images from the largest public available dataset with pixel accurate ground truths provided by [12], and compare with ten state-of-the-art methods: IT98 [5], MA03 [6], LC06 [15], GB07 [8], AC08 [7], FT09 [12], MSS10 [16], HC11 [17], HFT12 [11], CIS13 [18]. Results of these methods are obtained by two main ways: i) results provided by [12] (IT98 [5], MA03 [6], GB07 [8], AC08 [7]) and [17] (LC06 [15], HC11 [17]), ii)running the authors' publicly available code(FT09 [12], MSS10 [16], HFT12 [11], CIS13 [18]).

3.1. Objective evaluation

Similar as previous methods, we firstly evaluate our method using precision-recall analysis. Binary saliency maps are generated from each method using thresholds in the range of 0 and 255. Then we compute the precision and recall values of each method using binary saliency maps and ground truths. The corresponding Precision-Recall curves are shown in Fig.3 (a). We can see that our method outperforms others methods. The recall value becomes minimum when the threshold is close to 255. As shown in Fig.4 (a), our minimum recall value is far higher than those of methods, which indicates that our method more precise to salient regions. However, it is insufficient to evaluate the performance of saliency models according to the limitation of precision recall analysis presented in [19].



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l) (m)

Fig.5 Visual comparison of saliency maps. (a) Original images, (b) Ground truth, (c)-(l) represent the results of IT98 [5], MA03 [6], LC06 [15], GB07 [8], AC08 [7], FT09 [12], MSS10 [16], HC11 [17], HFT12 [11], CIS13 [18], (m) Our results.

For a more comprehensive and balanced evaluation, we use two measures the area under the receiver operating characteristics curve (ROC Area) [11] and the mean absolute error (MAE) [19]. Our method efficiently outperforms other methods, while it only achieves similar performance as HC11 [17] in terms of ROC Area (see Fig.4 (b)). However, Fig.4 (c) illustrates that our results at least reduce the MAE by 37% compared to results of HC11 [17], which indicates our results has more similarities with ground truths and better suits for the quality of the weighted in some applications.

3.2. Visual comparison

Fig.5 demonstrates some examples of visual comparison of different methods. From Fig.5, we can see that most methods may fail to distinguish salient regions and backgrounds, when salient regions and backgrounds have similar color or backgrounds have complex contents. In contrast, our method can not only detect salient objects accurately compared with ground truth, but also uniformly highlight salient areas rather than only boundaries. Moreover, we can extract well-defined boundaries of salient objects.

Experiments testify to the superiority of our method over other methods in terms of objective evaluation and visual comparison. However, our method may fail to detect salient regions, since our result is unavoidably affected by the accuracy of FT method [12]. In addition, some parts of backgrounds with appearance divergence to the image boundaries may be incorrectly highlighted as salient regions. Fig.6 shows some failure cases.

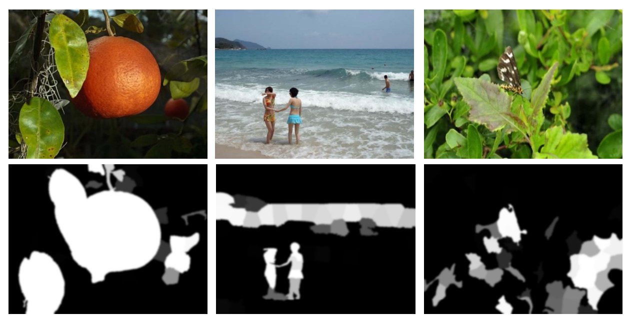


Fig.6 Failure examples

4. CONCLUSION

In this paper, we propose a saliency detection method based on a background probability model, which incorporates background priors and probability theory to measure saliency. Experimental results show that our method outperforms ten state-of-the-art methods and is simple to implement. However, this kind of approaches may fail if the superpixels on the image boundary cannot well represent the non-salient portion of the image. In the future, we will jointly consider color feature and other features to construct a better background model that contributes to obtain more precision saliency results.

REFERENCES

[1] D. Walther, U. Rutishauser, C. Koch, and P. Perona, “Selective visual attention enables learning and recognition of multiple objects in

cluttered scenes," *Computer. Vision. Image Understanding.*, vol.100, no. 1-2, pp. 41–63, Oct.-Nov. 2005.
[2] G.-X. Zhang, M.-M. Cheng, S.-M. Hu, and R. R. Martin. "A shape-preserving approach to image resizing," *Computer Graphics Forum*, 28(7):pp.1897–1906, 2009.
[3] M. Donoser, M. Urschler, M. Hirzer, and H. Bischof. "Saliency driven total variation segmentation," *in Proc. IEEE International Conference on Computer. Vision*, pp.817–824, 2009.
[4] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Trans. Image Process.*, vol. 19, no.1, pp. 185–198, Jan. 2010.
[5] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pp. 1254–1259, Nov. 1998.
[6] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," *in Proc. ACM International Conference on Multimedia*, pp. 374–381, November 2003.
[7] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk, "Salient region detection and segmentation," *International Conference on Computer Vision Systems*, vol. 5008, pp. 66–75, 2008.
[8] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Advances in Neural Information Processing Systems*, pp. 545–552, 2007.
[9] S. Frintrop, M. Klodt, and E. Rome, "A real-time visual attention system using integral images," *International Conference on Computer Vision Systems*, March 2007.
[10] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," *IEEE Conference on Computer Vision and Pattern Recognition,* pp. 1–8, June 2007.
[11] J. Li, M. D. Levine, X. J. An, X. Xu, and H. G. He, "Visual saliency based on scale-space analysis in the frequency domain," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.35, no.4, pp.996–1010, 2012.
[12] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk, "Frequency-tuned salient region detection," *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1597–1604, June 2009.
[13] B. Jiang, L.H. Zhang, H.C. Lu, C. Yang, and M. H. Yang, "Saliency detection via absorbing markov chain," *in Proc. IEEE International Conference on Computer. Vision*, pp.1665–1672, 2013.
[14] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk, "Slic superpixels," *Technical Report, EPFL,* 2010.
[15] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," *in Proc. ACM International Conference on Multimedia*, pp. 815–824, 2006.
[16] R. Achanta and S. S ¨ usstrunk, "Saliency detection using maximum symmetric surround", *in Proc. IEEE International Conference on Image Processing*, 2010.
[17] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," *in Proc. IEEE Conference on Computer Vision and Pattern Recognition,* pp. 409–416, 2011.
[18] S.Ryu, B. Ham, and K. Sohn, "Contextual information based visual saliency model," *in Proc. IEEE International Conference on Image Processing*, pp. 201–205, 2010.
[19] M.-M. Cheng, J. Warrell, W. Y. Lin, S. Zheng, V. Vineet, N. Crook, "Efficient salient region detection with soft image abstraction," *in Proc. IEEE Conference on Computer Vision and Pattern Recognition,* pp. 1529–1536, 2013.