

GEODESIC WEIGHTED BAYESIAN MODEL FOR SALIENT OBJECT DETECTION

Xiang Wang Huimin Ma Xiaozhi Chen

Tsinghua National Laboratory for Information Science and Technology(TNList)
Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
{wangxiang14, chenxz12}@mails.tsinghua.edu.cn, mhmpub@tsinghua.edu.cn

ABSTRACT

In recent years, a variety of salient object detection methods under Bayesian framework have been proposed and many achieved state of the art. However, those ignore spatial relationships and thus background regions similar to the objects are also highlighted. In this paper, we propose a novel geodesic weighted Bayesian model to address this issue. We consider spatial relationships by attaching more importance to regions which are more likely to be parts of a salient object, thus suppressing background regions. First, we learn a combined similarity via multiple features to measure similarity of adjacent regions. Then, we apply the combined similarity as edge weight to construct an undirected weighted graph and compute geodesic distance. Last, we utilize the geodesic distance to weight the observation likelihood to infer a more precise saliency map. Experiments on several benchmark datasets demonstrate the effectiveness of our model.

Index Terms— Salient object detection, geodesic weight, Bayesian framework, superpixel

1. INTRODUCTION

Salient object detection has been an important research area in computer vision, with its wide applications such as image segmentation [1], object recognition [2] and content-aware image editing [3]. Due to the lack of a clear definition, almost all bottom-up methods try to compute a saliency map using some assumptions on object and background. Center-surround difference [4, 5] is an early and widely used metric which assumes that the contrast between salient objects and their surrounding regions is high. In addition, center prior [6, 7] and backgroundness prior [8] have been proposed based on the observation that objects usually lie near the center of the image and thus the border of an image is more likely to be background. However, these bottom-up methods suffer from two main drawbacks. First, these bottom-up methods are sensitive to noise when the background is complex. Second, they are unable to uniformly highlight the whole salient object.

The Bayesian framework is an effective model to address the above problems. However, a general Bayesian framework can only make small improvement since it ignores spatial relationships. As a result, background regions similar to object will also be highlighted. In [9], Bayesian model using low and mid level cues was proposed and achieved state-of-the-art result. However, the main reason for this was its time-consuming computation for prior distribution, and it also showed a weakness in suppressing background regions effectively.

Motivated by the problems with bottom-up methods, and the drawbacks of existing Bayesian framework mentioned above, we propose an effective Geodesic Weighted Bayesian model to improve the quality of salient object detection. We consider spatial relationships by attaching more importance to regions which are more likely to be parts of a salient object, thus suppressing the background regions. Geodesic distance is an effective metric which considers both appearance similarities and spatial distance, so we utilize geodesic distance as the weight. Our main contributions are threefold. First, we propose an effective method for extracting initial salient regions. Second, we learn a combined similarity via multiple feature to measure the similarity of adjacent regions, namely, the probability of being parts of the same object. Third, we utilize the combined similarity as edge weight to construct an undirected weighted graph to compute geodesic distance. The geodesic distance is used to weight observation likelihood. Using the saliency maps of existing methods as prior distribution, we obtain a geodesic weighted Bayesian model which generates more precise saliency maps. Our model can be integrated to all existing methods and improve the quality of most methods.

2. THE PROPOSED METHOD

Given an image, we segment it into superpixels (*i.e.*, regions, used interchangeably) using the SLIC algorithm [10]. Then as in [9], the Bayesian inference for estimating saliency map is formulated as

$$p(sal|\mathbf{v}) = \frac{p(sal)p(\mathbf{v}|sal)}{p(sal)p(\mathbf{v}|sal) + p(bg)p(\mathbf{v}|bg)}, \quad (1)$$

This work was supported by National Natural Science Foundation of China (No. 61171113).

where $p(sal)$ denotes the prior distribution of the salient regions and $p(bg) = 1 - p(sal)$. \mathbf{v} denotes the feature vector of a given pixel. $p(\mathbf{v}|sal)$ and $p(\mathbf{v}|bg)$ (shorthand for $p(\mathbf{v}|sal = 1)$ and $p(\mathbf{v}|bg = 1)$) denote the observation likelihoods which are computed inside and outside the initial salient regions, respectively.

2.1. Extracting Initial Salient Regions

In contrast to [9] which computes a convex hull via interest points, we initialize the salient regions using binarized region contrast weighted by spatial distance [11] based on our observation that region contrast can highlight the most salient regions effectively in most cases. A superpixel's contrast is formulated as

$$Ctr(p) = \sum_{i=1}^N d_{sp}(p, p_i) w_{pos}(p, p_i), \quad (2)$$

where $d_{sp}(p, p_i) = \max(colDist(p, p_i), lbpDist(p, p_i))$ denotes the feature distance between superpixel p and p_i , with $colDist(p, p_i)$ and $lbpDist(p, p_i)$ representing the χ^2 distance of Lab color histograms and Local Binary Patterns (LBP) [12, 13] histograms, respectively. $w_{pos}(p, p_i) = \exp(-\frac{d_{pos}^2(p, p_i)}{2\sigma_{pos}^2})$ denotes the spatial weight, while $d_{pos}(p, p_i)$ denotes the Euclidean distance between the center of superpixel p and p_i , and $\sigma_{pos} = 0.25$ as in [14, 15]. N represents the number of superpixels. The initial salient region set $SalSPs$ is defined as

$$SalSPs = \{p | Ctr(p) > th\}, \quad (3)$$

with $th = 0.75$ which was empirically found to give.

2.2. Geodesic Weight

The general Bayesian framework often suffers from highlighting regions in background which are similar to objects since it ignores spatial relationships. To address this issue, we attach more importance to regions which are not only similar but also near to objects, namely, regions which are more likely to be parts of a salient object. Geodesic distance is an effective metric which considers both appearance similarities and spatial distance. To compute the geodesic distance between regions, we construct an undirected weighted graph by connecting all adjacent superpixels. To measure the similarity of adjacent regions, namely, the probability of being parts of the same object, we learn a combined similarity via multiple features and use it as edge weight of the graph. The combined similarity $\Phi(p_i, p_j)$ between superpixel p_i and p_j is formulated as

$$\Phi(p_i, p_j) = \sum_{s \in \{c, t, b\}} w_s f_s(p_i, p_j) + b, \quad (4)$$

with w_s denotes the feature weight and b denotes the bias.

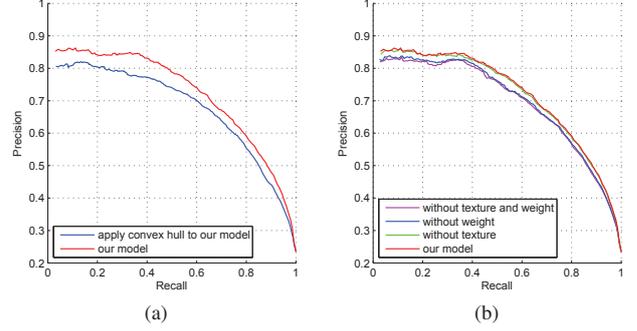


Fig. 1. (a) A comparison between convex hull in [9] and our initial salient regions (Sec. 2.1). (b) A comparison between three incomplete models and our model. We use the saliency map of FT [17] as prior distribution for example, similar results are also observed on other methods but omitted here for brevity.

Table 1. Feature weights

Feature	Color Similarity	Texture Similarity	Common Border Ratio	b
Weight	8.99	2.25	1.40	-3.28

The feature similarities we used are:

Color Similarity f_c : color similarity is an effective cue to measure the similarity between two regions. We define color similarity using χ^2 distance of Lab color histograms with 32 bins for each channel.

Texture Similarity f_t : texture similarity is complementary to color similarity. We define the texture similarity using χ^2 distance of LBP histograms with 32 bins for each region.

Common Border Ratio f_b [16]: the common border ratio represents the connections between two adjacent regions which is formulated as $f_b(i, j) = \max(\frac{l_{ij}}{l_i}, \frac{l_{ij}}{l_j})$,

where l_i and l_j represent the perimeters of superpixels i and j , respectively, l_{ij} represents the length of their common border.

The weights we have learned via SVM are shown in Table 1. Then we normalize the combined similarity using sigmoid function as

$$Sim(p_i, p_j) = \frac{1}{1 + \exp(-\Phi(p_i, p_j))}. \quad (5)$$

So the edge weight $W_e(p_i, p_j)$ between vertex p_i and p_j is

$$W_e(p_i, p_j) = 1 - Sim(p_i, p_j). \quad (6)$$

The geodesic distance d_{geo} is defined as the accumulated edge weights along their shortest path on the graph [15], then the geodesic weight is defined as

$$W_{geo}(p_i, p_j) = \exp(-\frac{d_{geo}^2(p_i, p_j)}{2\sigma_{geo}^2}). \quad (7)$$

Empirically, we set $\sigma_{geo} = 0.1$ in our experiments.

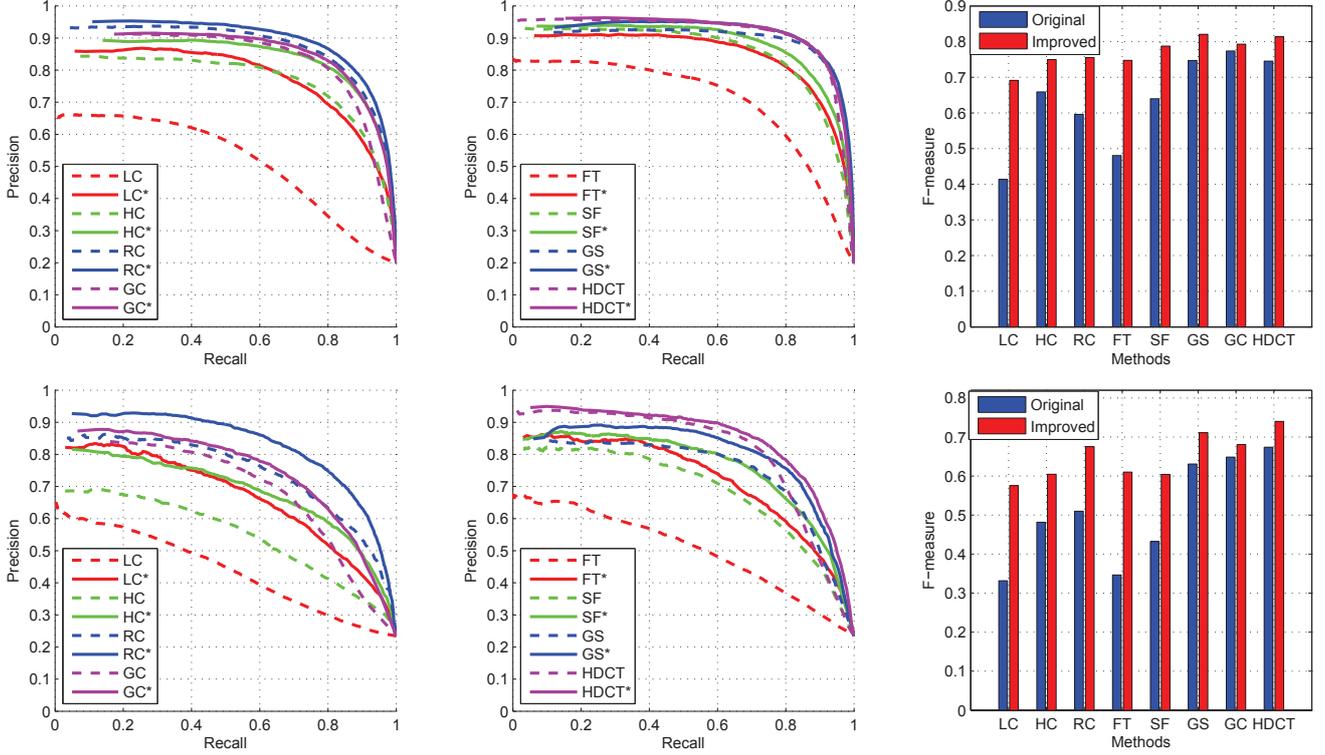


Fig. 2. Comparison of different methods with their improved versions (*). The first row are tested on ASD [17] and the second are on CSSD [18]. The first two columns show the improvement of PR curves and the last column shows the improvement of F-measure.

2.3. Geodesic Weighted Observation Likelihood

We consider spatial relationships by attaching more importance to the regions near (in geodesic distance) to the initial salient regions, the observation likelihood $p(\mathbf{v}|sal)$ and $p(\mathbf{v}|bg)$ is formulated by regions and weighted by the geodesic weight as

$$p(\mathbf{v}|sal) = \sum_{p_i \in SalSPs} p_{geo}^{(sal)}(p_i) p(\mathbf{v}|sal, p_i), \quad (8)$$

$$p(\mathbf{v}|bg) = \sum_{p_i \notin SalSPs} p_{geo}^{(bg)}(p_i) p(\mathbf{v}|bg, p_i), \quad (9)$$

with

$$p_{geo}^{(sal)}(p_i) = \frac{\text{mean}(W_{geo}(p_i, SalSPs))}{\sum_{p_j \in SalSPs} \text{mean}(W_{geo}(p_j, SalSPs))}, \quad (10)$$

$$p_{geo}^{(bg)}(p_i) = \frac{\text{mean}(W_{geo}(p_i, SalSPs))}{\sum_{p_j \notin SalSPs} \text{mean}(W_{geo}(p_j, SalSPs))}, \quad (11)$$

which denote the normalized mean geodesic distance between superpixel p_i and the initial salient superpixels.

In [9], given a pixel x , the feature vector \mathbf{v} is only represented by its Lab color channels. In our work, we also adopt

texture feature LBP, *i.e.*, $\mathbf{v}(x) = [l(x), a(x), b(x), lbp(x)]$. Then the observation likelihood of a given pixel x in superpixel p_i is calculated similar as [19, 9],

$$p(\mathbf{v}|sal, p_i) = \prod_{f \in \{l, a, b, lbp\}} \frac{N_{p_i(f(x))}}{N_{p_i}}, \quad p_i \in SalSPs, \quad (12)$$

$$p(\mathbf{v}|bg, p_i) = \prod_{f \in \{l, a, b, lbp\}} \frac{N_{p_i(f(x))}}{N_{p_i}}, \quad p_i \notin SalSPs, \quad (13)$$

where N_{p_i} is the number of pixels within superpixel p_i , $N_{p_i(f(x))}$ is the number that superpixel p_i contains $f(x)$. $f \in \{l, a, b, lbp\}$ denotes the component of feature vector \mathbf{v} .

In summary, substituting observation likelihood (8) and (9) into (1), and utilizing saliency map of existing methods as prior distribution, we obtain a geodesic weighted Bayesian model which generates a more precise saliency map.

3. EXPERIMENTS

We test our method on two standard benchmark datasets: ASD [17] and CSSD [18]. ASD which contains 1000 images is widely used and relatively simple while CSSD containing 200 images which is more challenging. The code of our proposed model is available on our project site.

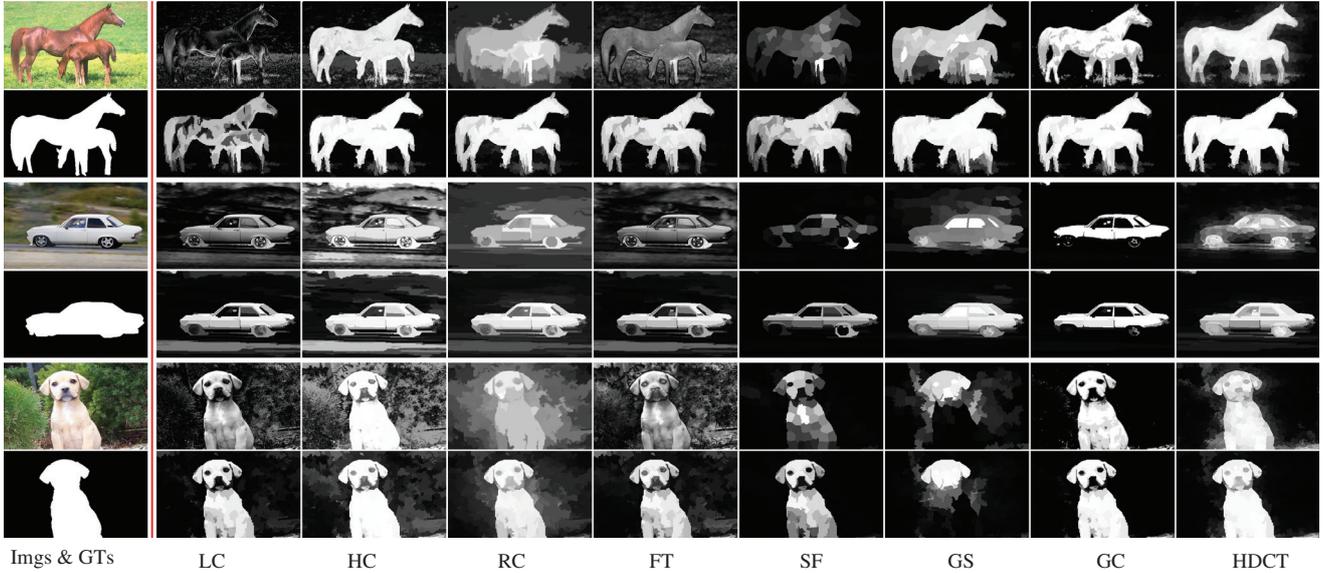


Fig. 3. Qualitative comparison of numerous methods with their improved versions. The first column shows source images and ground truth. For the second to the last column, the odd rows show saliency maps of existing methods, the even rows show their improved results via our model. The background regions are strongly suppressed and the improved maps are more uniform.

3.1. Evaluation of the Effectiveness of Our Model

We extensively conduct experiments to verify the effectiveness of our model. To evaluate our proposed initial salient regions in Sec. 2.1, we apply the convex hull in [9] to our model and compare their performances. To evaluate our proposed texture feature in feature vector and geodesic weight in computing observation likelihood in Sec. 2.3, we remove them from our model respectively, thus we get three incomplete models: model without texture feature and geodesic weight, model without texture feature, and model without geodesic weight. The results are shown in Fig.1. Fig.1(a) demonstrates that our method for extracting initial salient region is more effective than the convex hull proposed in [9] in improving the quality of saliency map. From Fig.1(b), we can conclude that both texture feature and geodesic weight make significant contributions to our model, and geodesic weight has a greater impact.

3.2. Comparison with State-of-the-Art Methods

We integrate our model into numerous state-of-the-art methods, namely, utilizing their saliency map as prior distribution: LC [20], FT [17], HC [11], RC [11], SF [14], GS [8], GC [21], HDCT [22].

For performance evaluation, we utilize precision-recall curves (PR curves) and F-measure. We normalize the saliency map to $[0, 255]$ and then binarize it with threshold from 0 to 255, so we get 256 pairs of precision-recall data, the PR curves are calculated by averaging them on each dataset. We

compute the F-measure using Eq.14 for each precision-recall pair and report the average,

$$F_{\beta} = \frac{(1 + \beta^2)Precision \times Recall}{\beta^2Precision + Recall}. \quad (14)$$

As suggested in many prior works, we set $\beta^2 = 0.3$.

We utilize the saliency map of the above methods as the prior distribution in our model, Fig.2 shows the PR curves and F-measure with the original and improved methods compared. We can see that both PR curves and F-measure are significantly improved. In addition, the PR curves also show that our improved versions have a higher minimum recall value compared with original methods, which means our improved model is able to detect more precise saliency maps than before. A qualitative comparison is also shown in Fig. 3.

4. CONCLUSION

In this paper, we propose a novel geodesic weighted Bayesian model to improve the quality of salient object detection. Motivated by the drawback of the general Bayesian framework that it ignores spatial relationships, we attach more importance to regions which are more likely to be parts of salient object, thus suppressing background regions similar to salient object. Experiments demonstrate that the method for extracting initial salient regions is effective. Both texture feature and geodesic weight make significant contributions to our model, and geodesic weight has a greater impact. When integrated into existing methods, our approach significantly improves performances.

5. REFERENCES

- [1] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," *ACM TOG*, vol. 23, no. 3, pp. 309–314, 2004.
- [2] Ueli Rutishauser, Dirk Walther, Christof Koch, and Pietro Perona, "Is bottom-up attention useful for object recognition?," in *CVPR*, 2004, vol. 2, pp. II–37.
- [3] Guo-Xin Zhang, Ming-Ming Cheng, Shi-Min Hu, and Ralph R Martin, "A shape-preserving approach to image resizing," in *Computer Graphics Forum*, 2009, vol. 28, pp. 1897–1906.
- [4] Laurent Itti, Christof Koch, and Ernst Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE TPAMI*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [5] Tie Liu, Jian Sun, Nan-Ning Zheng, Xiaoou Tang, and Heung-Yeung Shum, "Learning to detect a salient object," in *CVPR*, 2007.
- [6] Huaizu Jiang, Jingdong Wang, Zejian Yuan, Tie Liu, Nanning Zheng, and Shipeng Li, "Automatic salient object segmentation based on context and shape prior," in *BMVC*, 2011, vol. 3, p. 7.
- [7] Risheng Liu, Junjie Cao, Zhouchen Lin, and Shiguang Shan, "Adaptive partial differential equation learning for visual saliency detection," in *CVPR*, 2014, pp. 3866–3873.
- [8] Yichen Wei, Fang Wen, Wangjiang Zhu, and Jian Sun, "Geodesic saliency using background priors," in *ECCV*, 2012, pp. 29–42.
- [9] Yulin Xie, Huchuan Lu, and Ming-Hsuan Yang, "Bayesian saliency via low and mid level cues," *IEEE TIP*, vol. 22, no. 5, pp. 1689–1698, 2013.
- [10] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE TPAMI*, vol. 34, no. 11, pp. 2274–2282, 2012.
- [11] Ming-Ming Cheng, Guo-Xin Zhang, Niloy J Mitra, Xiaolei Huang, and Shi-Min Hu, "Global contrast based salient region detection," in *CVPR*, 2011, pp. 409–416.
- [12] Timo Ojala, Matti Pietikainen, and David Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," *Pattern Recognition*, vol. 1, pp. 582–585, 1994.
- [13] Timo Ojala, Matti Pietikäinen, and David Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.
- [14] Federico Perazzi, Philipp Krahenbuhl, Yael Pritch, and Alexander Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *CVPR*, 2012, pp. 733–740.
- [15] Wangjiang Zhu, Shuang Liang, Yichen Wei, and Jian Sun, "Saliency optimization from robust background detection," in *CVPR*, 2014, pp. 2814–2821.
- [16] Santiago Manen, Matthieu Guillaumin, and Luc Van Gool, "Prime object proposals with randomized prim's algorithm," in *ICCV*, 2013, pp. 2536–2543.
- [17] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk, "Frequency-tuned salient region detection," in *CVPR*, 2009, pp. 1597–1604.
- [18] Qiong Yan, Li Xu, Jianping Shi, and Jiaya Jia, "Hierarchical saliency detection," in *CVPR*, 2013, pp. 1155–1162.
- [19] Esa Rahtu, Juho Kannala, Mikko Salo, and Janne Heikkilä, "Segmenting salient objects from images and videos," in *ECCV*, 2010, pp. 366–379.
- [20] Yun Zhai and Mubarak Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *ACM Multimedia*, 2006, pp. 815–824.
- [21] Ming-Ming Cheng, Jonathan Warrell, Wen-Yan Lin, Shuai Zheng, Vibhav Vineet, and Nigel Crook, "Efficient salient region detection with soft image abstraction," in *ICCV*, 2013, pp. 1529–1536.
- [22] Jiwhan Kim, Dongyoon Han, Yu-Wing Tai, and Junmo Kim, "Salient region detection via high-dimensional color transform," in *CVPR*, 2014.