



Geodesic weighted Bayesian model for saliency optimization[☆]



Xiang Wang^{a,b}, Huimin Ma^{a,b,*}, Xiaozhi Chen^{a,b}

^a Tsinghua National Laboratory for Information Science and Technology (TNList), China

^b Department of Electronic Engineering, Tsinghua University, China

ARTICLE INFO

Article history:

Received 23 September 2015

Available online 26 February 2016

Keywords:

Salient object detection

Saliency optimization

Bayesian framework

Geodesic weight

ABSTRACT

Bottom-up methods and general Bayesian framework for saliency detection commonly suffer from two drawbacks. First, they are sensitive to background noise, thus background regions similar to objects are also highlighted. Second, they only consider appearance features and thus object with several different parts will not be highlighted uniformly. In this paper, we propose a novel and unified geodesic weighted Bayesian model which considers spatial relationship by reformulating Bayes' formula. First, we infer a more precise initial salient regions via fully connected CRF model. Second, to highlight the whole object uniformly, we learn a robust measure of region similarity which describes the probability of two regions belonging to the same object, so regions belonging to the same object will be given similar saliency value. Third, using our learnt region similarity as edge weight, we construct an undirected weighted graph to compute geodesic distance of regions. Regions with short geodesic distance from initial salient regions will be attached more importance, thus suppressing background noise. By using results of existing methods as prior distribution, our model can integrate into all methods and improve their performance. Experiments on benchmark datasets demonstrate that our model can significantly improve the quality of saliency detection.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Salient object detection has been an important research field in computer vision, with its wide applications such as image segmentation [24], object recognition [25] and object detection [5]. Since saliency detection is a class-agnostic task and it lacks a clear definition, almost all bottom-up methods try to compute saliency map using some assumptions on salient object and background, such as center-surround difference [8,17], center prior [10,16] and backgroundness prior [27]. These assumptions are based on the observation that salient objects usually have high contrast with other regions and they mostly lie on the center of an image.

However, these methods suffer from two main drawbacks. First, they are sensitive to background noise when the background is complex. Second, they are unable to uniformly highlight the whole salient object when the object has some parts looking differently. Some examples are shown in Fig. 1.

The Bayesian framework is an effective model to address the above problems. Xie et al. [28] proposed a Bayesian model using low and mid-level cues which achieved state-of-the-art result. However, it also showed weaknesses in the above two issues since general Bayesian framework only considers the appearance feature while the spatial relationship is ignored.

In this paper, we propose a novel geodesic weighted Bayesian model (GWB) which considers spatial relationship by attaching more importance to regions which are more likely to be parts of salient objects. Our main contributions are threefold. First, we propose a fully connected CRF model to infer more precise initial salient regions, thus making the computation of observation likelihood in the Bayesian framework more accurate. Second, to highlight the whole object uniformly, we learn a robust region similarity to measure the probability of two regions belonging to the same object. With that, regions that have high probability to be one object will be given similar saliency value. Third, we construct an undirected weighted graph using our proposed region similarity as edge weight to compute geodesic distance. Then regions with short geodesic distance from the initial salient regions will be attached more importance. Thus the object regions are highlighted and the background is suppressed.

In contrast to [28] which computes prior distribution via time-consuming superpixel clustering, we use saliency maps of existing

[☆] This paper has been recommended for acceptance by M. Couprie.

* Corresponding author at: Department of Electronic Engineering, Tsinghua University, China. Tel.: +86 10 62781432; fax: +86 10 62770317.

E-mail addresses: wangxiang14@mails.tsinghua.edu.cn (X. Wang), mhmpub@tsinghua.edu.cn, mhmpub@gmail.com (H. Ma), chenxz12@mails.tsinghua.edu.cn (X. Chen).

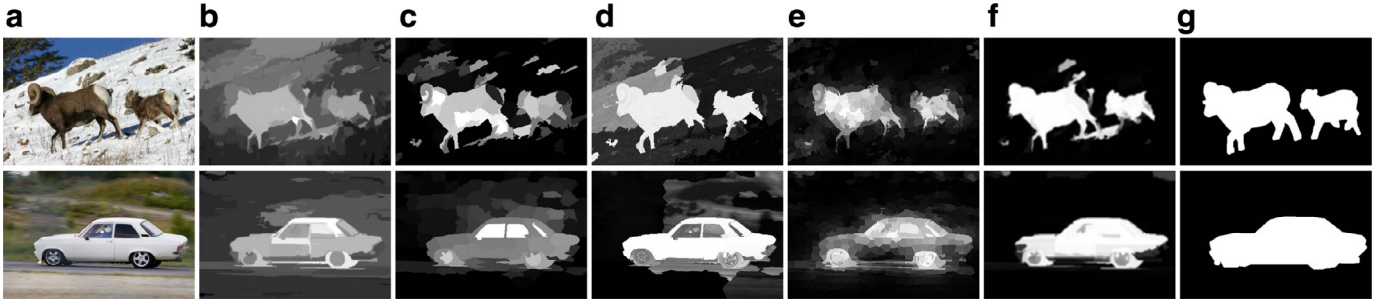


Fig. 1. Some examples. (a) Source images, saliency maps generated by (b) Cheng et al. [6], (c) Wei et al. [27], (d) Xie et al. [28], (e) Kim et al. [11], (f) our proposed method, and (g) ground truth. Compared with these methods, our results are more uniform and the background noise is strongly suppressed.

methods as prior distribution, so our model can be integrated into existing methods and improve their performance.

This paper provides a more complete understanding of the GWB model first presented in the previous conference version [26]. We improve the previous framework in two aspects. First, instead of using binarized region contrast map as initial salient regions, we propose a fully connected CRF model to infer more precise initial salient regions. Second, a two-layer neighborhood graph is constructed to handle the case of occlusion in computing geodesic distance. Further details on background, motivation, method, analysis and evaluation are also given. Experiments on a broader range of datasets are conducted which demonstrate that our model effectively improve existing methods to state-of-the-art performance level, even though the original saliency maps are not satisfying.

The rest of this paper is organized as follows. Section 2 introduces and compares with related work, Section 3 introduces the details of the proposed model, Section 4 presents the experiments and conclusion is made in Section 5.

2. Related work

CRF/MRF is an energy minimum model which has been widely applied in semantic segmentation and salient object detection [3,4,7,9,12,13,23,31]. According to their purposes, these methods can be classified into two categories: segmentation [13,23] and estimation [3,4,7,9,12,31].

Rahtu et al. [23] proposed to segment salient object from images and videos using CRF with unary saliency term and color term. While traditional adjacency CRF models only consider neighboring nodes in the pairwise term, fully connected CRF can represent long-range interactions and thus making the segmentation results more accurate. However, the computation for inferring fully connected CRF is very expensive. Krähenbühl and Koltun [13] proposed an efficient inference algorithm which makes the fully connected CRF model practical. They defined the fully connected CRF on pixels and applied it to semantic segmentation. In this paper, we propose a fully connected CRF model to segment initial salient regions for computing likelihoods in Bayesian framework (Section 3.2) similar with [13], but we define the CRF model on superpixels considering that pixels in the same superpixel mostly share the same saliency value, we also utilize more features which are complementary. With that, we can segment more accurate initial salient region with high efficiency (33 FPS, see Section 4.1).

On the other hand, CRF/MRF model can be used for saliency estimation. Kocak et al. [12] and Yang et al. [31] proposed to jointly learn category-specific dictionary and CRF to localize salient objects with specific category. Jia and Han [9] proposed to use objectness score as coarse saliency map and then refine it via MRF by adding two abstract nodes which are connected with all other nodes to get final saliency map. Fu et al. [7] proposed to smooth spatial-temporal saliency map in videos via CRF by constructing a

two-frame graph. Aytekin et al. [3,4] proposed quantum cuts for salient object detection by linking quantum mechanics with spectral graph clustering.

Bayesian Framework is a probability formula which computes posterior distribution via prior distribution and observation likelihood. Rahtu et al. [23] proposed to measure saliency by applying Bayes' formula in sliding windows and comparing pixels inside and outside the rectangular inner window. Li et al. [15] proposed to estimate saliency map via dense and sparse reconstruction and then integrate these two maps via Bayesian framework by taking one saliency map as prior distribution and using the other one as likelihood. However, they got initial salient regions by binarizing the prior map with a threshold, which is less accurate and may bring bias since the initial salient regions are strongly correlated with the prior distribution. The closest work to ours is [28]. In [28], the initial salient regions is defined as a convex hull formed by interest points, and the prior distribution is defined as the overlap between convex hull and superpixel clustering partitions. We differ from that in three aspects. First, we propose a fully connected CRF model to infer initial salient regions which are more accurate than that in [28]. Second, we apply Bayesian framework as a unified optimization framework which takes existing saliency maps as prior distribution while Xie et al. computed prior distribution by time-consuming superpixel clustering method. Third, general Bayesian framework suffers from highlighting background regions with similar appearance features with salient object. We address this issue by considering spatial relationship in formulating observation likelihood and thus the background regions are strongly suppressed and the salient objects are highlighted uniformly.

3. Approach

3.1. Overview

Given an image, we first segment it into superpixels (i.e., regions, used interchangeably in this paper) using the SLIC algorithm [2]. Then as in [28], the Bayesian inference for estimating saliency map is formulated as:

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

where $p(sal)$ and $p(bg) = 1 - p(sal)$ denote the prior distribution of the salient regions and background, respectively. \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 likelihood which are computed inside and outside the initial salient regions, respectively. In our work, the initial salient regions are inferred using a fully connected CRF model (Section 3.2). We take saliency maps of existing methods as prior distribution $p(sal)$ and compute posterior distribution $p(sal|\mathbf{v})$ via Bayes' formula as the improved version of the existing maps. So our method is a unified

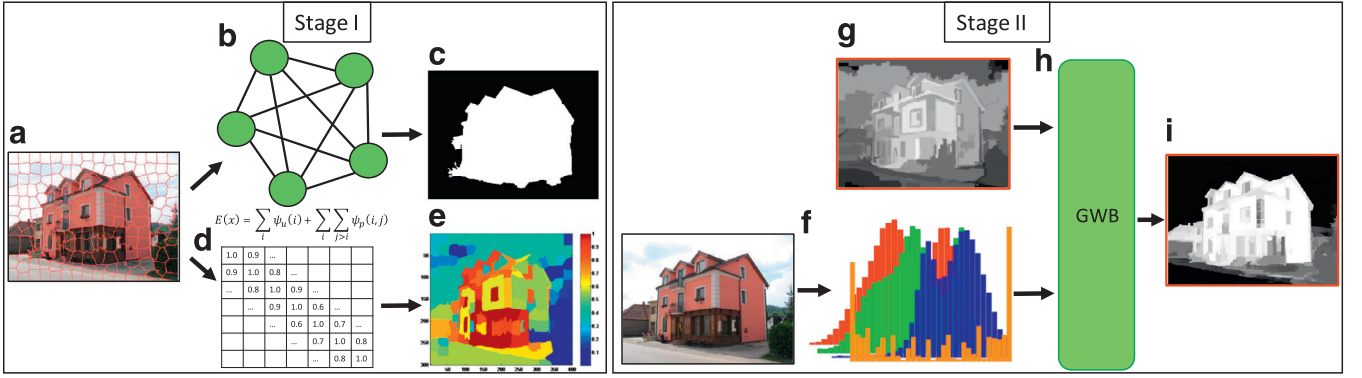


Fig. 2. Overview of the proposed approach. At stage I, we formulate the GWB model, and at stage II, we improve saliency map of existing methods via GWB to get a better map. (a) Segment source image into superpixels, (b) fully connected CRF model for inferring (c) initial salient regions, (d) compute geodesic distance using our learnt region similarity as edge weight and (e) form geodesic weight which is defined as normalized mean geodesic distance from initial salient regions. So far, we get geodesic weighted Bayesian model (GWB) by reformulating Bayes' formula (Eqs. (1)–(3)). At stage II, we compute (f) LAB color feature and LBP texture feature at each pixel as feature vector \mathbf{v} and use (g) saliency map of existing methods as prior distribution to infer a more precise saliency map (i) via (h) GWB.

optimization framework which only takes prior maps and source images as input.

While the general Bayesian framework suffers from highlighting background regions similar to the salient objects, we consider spatial relationship by attaching more importance to regions which are more likely to be parts of a salient object, thus suppressing background regions. Geodesic distance is an effective metric which considers both appearance similarities and spatial distance, so we utilize geodesic distance as the weight by splitting up the formulation of the observation likelihood $p(\mathbf{v}|\text{sal})$ and $p(\mathbf{v}|\text{bg})$. We formulate $p(\mathbf{v}|\text{sal})$ as:

$$\begin{aligned} p(\mathbf{v}|\text{sal}) &= \sum_{s_i \in I} p_{\text{geo}}(s_i|\text{sal}) p(\mathbf{v}|\text{sal} \cap s_i) \\ &= \sum_{s_i \in \text{sal}} p_{\text{geo}}(s_i|\text{sal}) p(\mathbf{v}|\text{sal} \cap s_i) \\ &= \sum_{s_i \in \text{sal}} p_{\text{geo}}(s_i|\text{sal}) p(\mathbf{v}|s_i), \end{aligned} \quad (2)$$

where s_i denotes superpixel i , I denotes the whole image, sal denotes initial salient regions, $p_{\text{geo}}(s_i)$ denotes the probability of s_i , namely, the weight of superpixel s_i . In our work, the weight is represented by normalized geodesic distance (Section 3.4). In the same way, $p(\mathbf{v}|\text{bg})$ is formulated similarly as:

$$p(\mathbf{v}|\text{bg}) = \sum_{s_i \in \text{bg}} p_{\text{geo}}(s_i|\text{bg}) p(\mathbf{v}|s_i), \quad (3)$$

where bg denotes initial background regions.

Fig. 2 shows the overview of our approach.

3.2. Initial salient regions

In contrast to [28] which detects coarse initial salient regions using convex hull via interest points, we aim to extract more precise regions thus making the observation likelihood $p(\mathbf{v}|\text{sal})$ and $p(\mathbf{v}|\text{bg})$ in (1) more accurate. In this paper, we formulate the extraction of initial salient regions as a binary segmentation problem. Fully connected CRF is an efficient model for image segmentation for its strong expressive power in long-range interactions, and with a highly efficient inference algorithm proposed in [13], we propose to extracting the initial salient regions using a fully connected CRF model. Since in most cases, pixels in the same superpixel share the same saliency value, we define the CRF model on the complete set of superpixels, this also saves computation time. The Gibbs energy

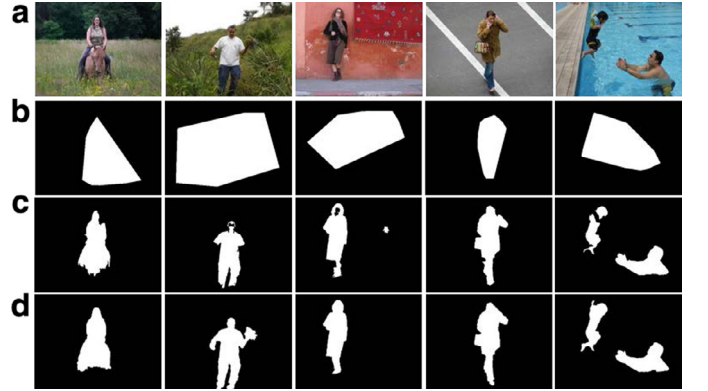


Fig. 3. (a) Source images. (b) Coarse initial salient regions in [28]. (c) Initial salient regions inferred by our CRF model. (d) Ground truth.

of our CRF model is defined as:

$$E(\mathbf{x}) = \sum_i \psi_u(x_i) + \sum_{i < j} \psi_p(x_i, x_j). \quad (4)$$

The unary term $\psi_u(x_i)$ measures the cost of assigning a label to superpixel i . In our work, we use the saliency map of existing method as the unary potential. The pairwise term $\psi_p(x_i, x_j)$ encourages similar and nearby superpixels to take similar labels which has the form of Potts model [22]:

$$\psi_p(x_i, x_j) = [x_i \neq x_j] k(\mathbf{f}_i, \mathbf{f}_j) \quad (5)$$

$$\begin{aligned} k(\mathbf{f}_i, \mathbf{f}_j) &= \omega_1 \exp\left(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|\mathbf{v}_i - \mathbf{v}_j|^2}{2\theta_\beta^2}\right) \\ &\quad + \omega_2 \exp\left(-\frac{|p_i - p_j|^2}{2\theta_\gamma^2}\right) \end{aligned} \quad (6)$$

where p_i denotes the position of the center of superpixel i , \mathbf{v}_i represents the feature vector of superpixel i which is concatenated by the mean LAB color and Local Binary Patterns (LBP) [19,20] histograms. All the parameters $\omega_1, \omega_2, \theta_\alpha, \theta_\beta, \theta_\gamma$ are learnt via cross validation. We use the highly efficient approximate inference algorithm proposed by Krähenbühl and Koltun [13] to infer the initial salient regions. Fig. 3 shows the results of the fully connected CRF model. We can see that our initial salient regions are very similar to ground truth.

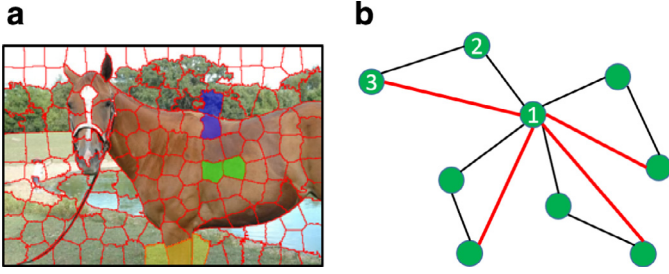


Fig. 4. (a) Example for selecting samples to train SVM. The green colored regions are a pair of positive samples and the blue colored regions are a pair of negative samples. Example for constructing two-layer neighborhood graph (yellow colored regions). The leg of the horse cuts off the connection between the two grass regions, with two-layer neighborhood graph, the disconnected grass regions become connected. Panel (b) shows an example. The black lines denote connections in general graph. Since node 1 and node 3 have the same neighbor: node 2, they are connected in two-layer neighborhood graph. Red lines show the additional connections of node 1 in two-layer neighborhood graph. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3. Learn a robust measure of region similarity

Our main idea is attaching more importance to regions which are more likely to be parts of a salient object. So we need to measure the region similarity, i.e., the probability of two regions belonging to the same object. Measuring the similarity between any two regions directly may be inefficient since only appearance feature can be considered while spatial relationship is ignored. So we propose to measure the similarity of adjacent regions first and then the similarity between any regions is defined as their geodesic distance.

Based on our observation and prior works [18], both appearance feature and connections between regions are important for measuring the probability of two regions belonging to the same object. So we use color similarity, texture similarity and common border ratio as basic features and combine them to get a robust measure.

The color similarity (f_c) and texture similarity (f_t) are two important and complementary metrics for measuring the appearance similarity. We define color similarity and texture similarity using χ^2 distance of LAB color histograms and LBP histograms with 32 bins for each region, respectively. The common border ratio (f_b) is another important measure which describes the connections between regions [18]. The common border ratio is defined as the maximum ratio between their common border and each of their perimeter:

$$f_b(i, j) = \max\left(\frac{l_{ij}}{l_i}, \frac{l_{ij}}{l_j}\right), \quad (7)$$

Table 1
Feature weights.

Feature	Color similarity	Texture similarity	Common border ratio	b
Weight	8.99	2.25	1.40	−3.28

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

We combine the above metrics to get a robust measure of region similarity and learn their weights using support vector machine (SVM):

$$\Phi(i, j) = \sum_{s \in \{c, t, b\}} w_s f_s(i, j) + b. \quad (8)$$

We select couples of adjacent regions within the ground truth as positive samples, and the negative samples are selected from adjacent regions with one region within the ground truth and another one within the background. Fig. 4 shows an example.

The weights we have learnt via SVM are shown in Table 1. We normalize the combined region similarity using a sigmoid function:

$$\text{Sim}(i, j) = \frac{1}{1 + \exp(-\Phi(i, j))}. \quad (9)$$

3.4. Geodesic weighted observation likelihood

To measure the similarity between any two regions, we build an undirected weighted graph on the whole set of superpixels with adjacent superpixels connected. The edge weight between adjacent nodes i and j is $W_e(i, j) = 1 - \text{Sim}(i, j)$.

We observed that occlusion often happens between objects and background. When occlusion happens, the connected regions of one object will be cut off by another object. Fig. 4 shows an example. So we construct a two-layer neighborhood graph, i.e., a graph in which nodes that have the same neighbors are also connected, to address this issue. In this way the disconnected regions become connected, and in addition, regions in the same object with long distance will get shorter geodesic distance, which makes the geodesic distance more accurate for measuring region similarity.

The geodesic distance d_{geo} is defined as the accumulated edge weights along their shortest path on the graph [33],

$$d_{geo}(i, j) = \min_{p_1=i, p_2, \dots, p_n=j} \sum_{k=1}^{n-1} W_e(p_k, p_{k+1}). \quad (10)$$

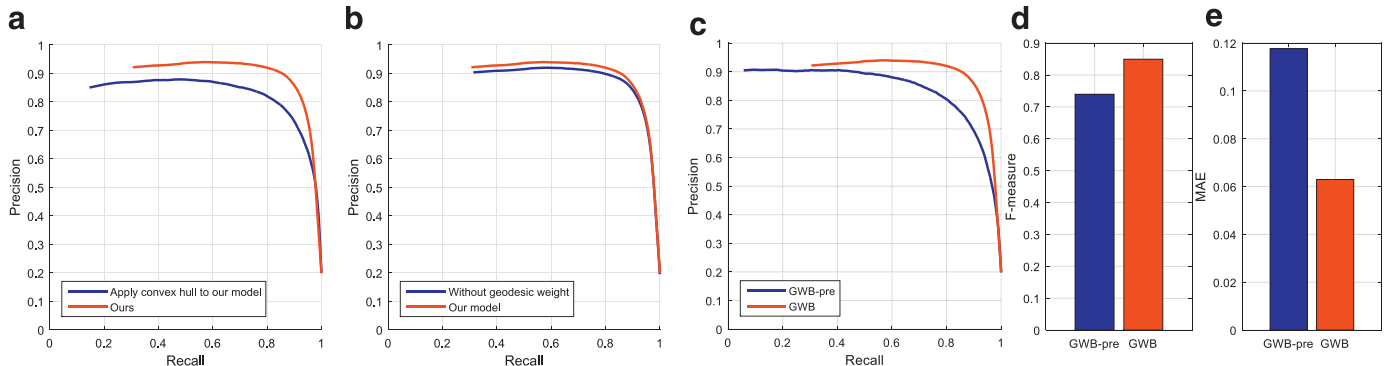


Fig. 5. (a) A comparison between convex hull in [28] and our initial salient regions. (b) Evaluating the effectiveness of our geodesic weight. (c–e) Comparison between GWB and the previous version (GWB-pre) in PR curves, F-measure and MAE. We use the saliency map of FT [1] on ASD dataset as prior distribution for example, similar results are also observed on other methods but omitted here for brevity.

Table 2
Comparison between one-layer and two-layer neighborhood graph on DUT-ORMON dataset.

Prior distribution	LC	HC	RC	FT	SF	GS	HDCT	LPS
One-layer neighborhood	0.5291	0.5084	0.5622	0.5489	0.5659	0.5703	0.5792	0.5893
Two-layer neighborhood	0.5340	0.5137	0.5642	0.5534	0.5717	0.5743	0.5824	0.5899

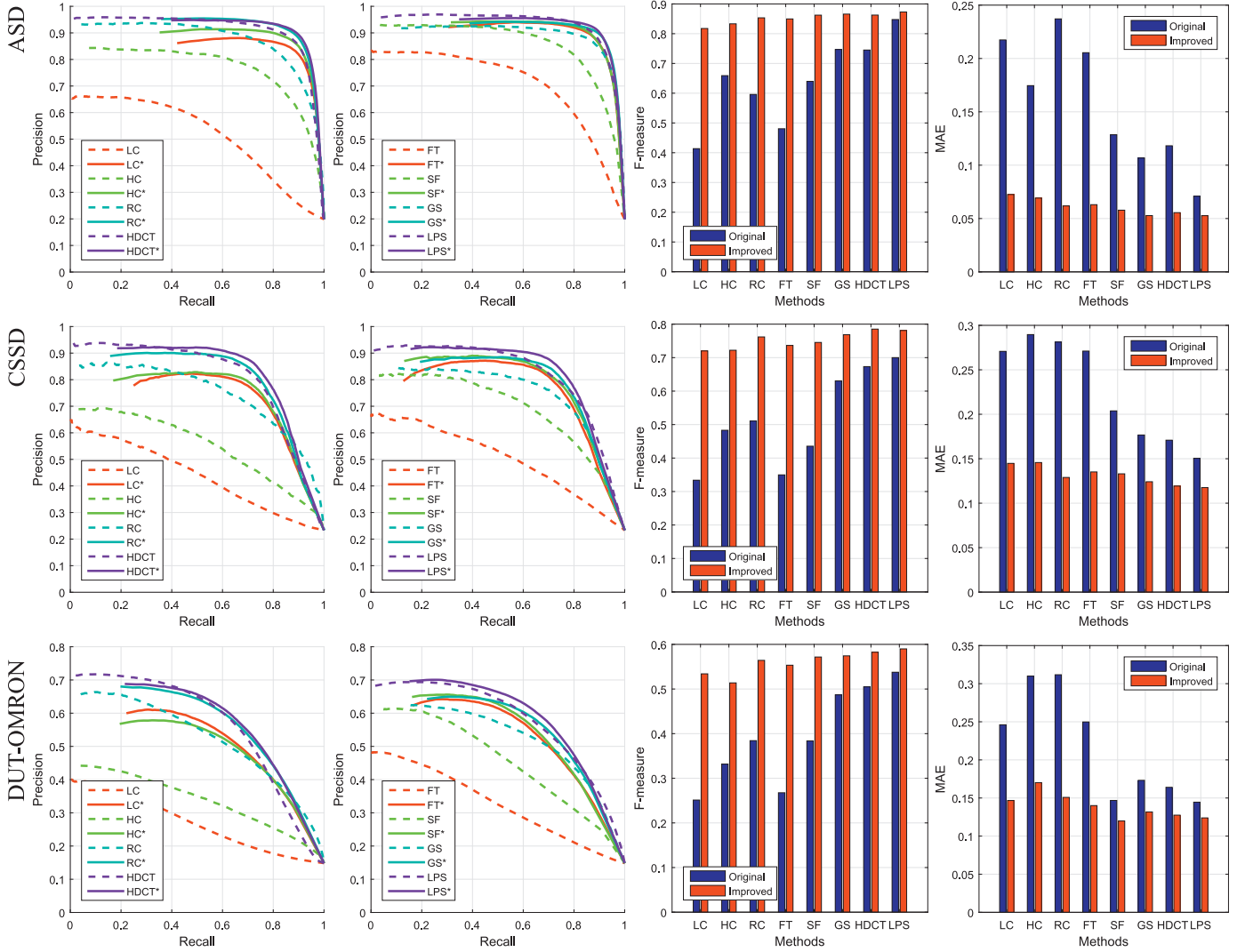


Fig. 6. Comparison of different methods with their improved versions (*). The first row are tested on ASD [1], the second are on CSSD [29], and the third are on DUT-ORMON [30]. The first two columns show the improvement of PR curves, the third column shows the improvement of F-measure, and the last column shows the decrease of mean absolute error (MAE).

We define the geodesic similarity as:

$$W_{geo}(i, j) = \exp\left(-\frac{d_{geo}^2(i, j)}{2\sigma_{geo}^2}\right). \quad (11)$$

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

The geodesic weight in (2) is formulated as normalized mean geodesic similarity between the given region and initial salient regions.

$$p_{geo}(s_i|sal) = \frac{\text{mean}(W_{geo}(s_i, sal))}{\sum_{s_j \in sal} \text{mean}(W_{geo}(s_j, sal))}, \quad (12)$$

$$p_{geo}(s_i|bg) = \frac{\text{mean}(W_{geo}(s_i, sal))}{\sum_{s_j \in bg} \text{mean}(W_{geo}(s_j, sal))}. \quad (13)$$

Given a pixel x , the feature vector is represented by its LAB color feature and LBP texture feature, i.e., $\mathbf{v}(x) = (l(x), a(x), b(x), lbp(x))$. The observation likelihood of a given pixel x in superpixel s_i in (2) and (3) is calculated as:

$$p(\mathbf{v}|s_i) = \prod_{f \in \{l, a, b, lbp\}} \frac{N_{s_i}(f(x))}{N_{s_i}}. \quad (14)$$

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

In summary, substituting (12)–(14) into (2) and (3) to compute observation likelihood, and then substituting them into (1), utilizing saliency map of existing methods as prior distribution, we obtain a geodesic weighted Bayesian model which generates a more precise saliency map.

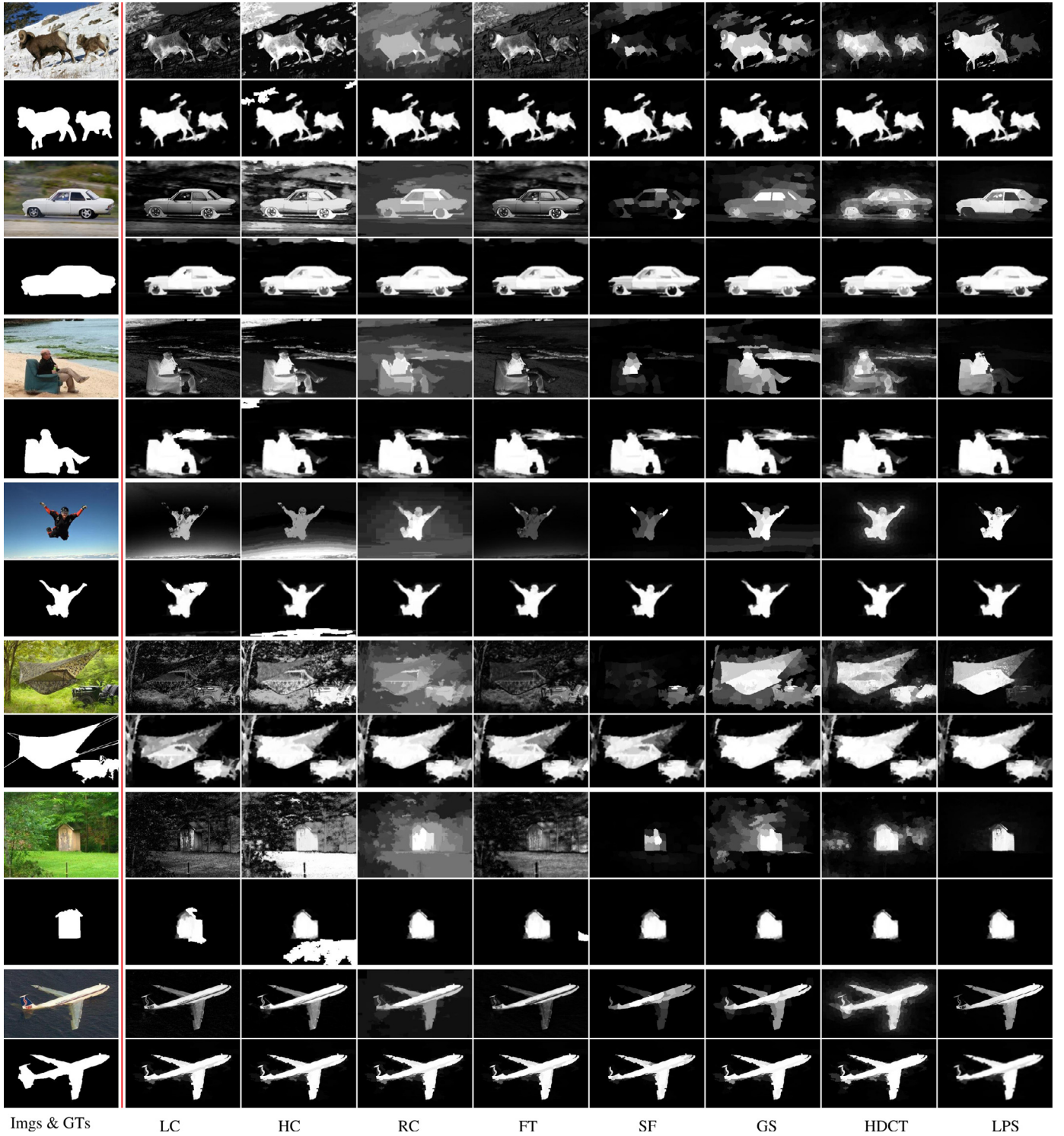


Fig. 7. Qualitative comparison of numerous methods with their improved versions. The first column shows source images and ground truth. For the second to the last column, the odd rows show saliency maps of existing methods, and the even rows show their improved results via our model.

4. Experiments

We test our method on three typical benchmark datasets: ASD [1], CSSD [29] and DUT-ORMON [30]. ASD which contains 1000 images is widely used and relatively simple, CSSD contains 200 images which are moderate difficult, DUT-ORMON contains 5168 images which are more challenging.

For performance evaluation, we utilize precision-recall curves (PR curves), F -measure and mean absolute error (MAE). The saliency maps are first normalized to $[0, 255]$ and then being binarized with a threshold sliding from 0 to 255. For each threshold, we compare the binary saliency map with ground truth and compute precision and recall for each image. The PR curves are then computed by averaging them on the whole dataset. F -measure is an overall performance measure which considers both precision

and recall:

$$F_\beta = \frac{(1 + \beta^2) \text{Precision} \times \text{Recall}}{\beta^2 \text{Precision} + \text{Recall}}, \quad (15)$$

as suggested in many prior works [1,6], we set $\beta^2 = 0.3$. Mean absolute error (MAE) describes the mean difference between continuous saliency map S (range from 0 to 1) and binary ground truth GT .

$$MAE = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H |S(x, y) - GT(x, y)| \quad (16)$$

4.1. Evaluation of the effectiveness of our model

We extensively conduct experiments to verify the effects of separate components and their combinations in our approach.

Initial salient regions: To evaluate our proposed initial salient regions in Section 3.2, we apply the convex hull in [28] to our model and compare their performances. Fig. 5(a) shows the comparison of precision-recall curves (PR curves). Our method significantly outperforms convex hull in [28], which means our initial salient regions are more precise and play an important role in our model.

Two-layer neighborhood graph: We compare our two-layer neighborhood graph in Section 3.4 with general one-layer neighborhood graph. Table 2 shows the results of F -measure on DUT-ORMON dataset when using saliency map of different methods as prior distribution. We can see that our two-layer neighborhood graph achieve a consistent improvement in F -measure with 0.4% on average.

Geodesic weight: To evaluate the effectiveness of our geodesic weight, we remove it from our model, i.e., assigning all weight equally, and compare the performance with the complete model. Fig. 5(b) shows the PR curves which demonstrate that geodesic weight makes significant contributions to our model.

Comparison with previous version: We compare our GWB model with the previous version GWB-pre and show some results in Fig. 5(c–e). Detailed comparison can be found in the Supplementary material. The results demonstrate that our GWB model improve our previous version by a large margin. In particular, the improved results via GWB have little dependency on the quality of original methods, namely, even though the original maps are not satisfying, the improved maps are also comparable to state-of-the-art, while the previous version GWB-pre has much dependency on the quality of original methods.

Speed: We evaluate our method on a 3.5 GHz CPU. The running time for GWB is 0.36 s, including 0.03 s for initial salient regions segmentation, 0.09 s for superpixel segmentation, 0.05 s for measuring region similarity, 0.14 s for computing all distributions and inference via Bayesian framework, and 0.05 s for other pre-processing, such as colorspace transformation and mean color and position of superpixels. Since most methods also include superpixel segmentation and same pre-processing, when integrating into existing methods, the additional running time of GWB can be further reduced. Thus our GWB brings little computational overhead to existing methods.

4.2. Integration and comparison with state-of-the-art methods

We integrate our model into numerous state-of-the-art methods, namely, taking their saliency map as prior distribution and inferring improved maps via our method: LC [32], FT [1], HC [6], SF [21], GS [27], HDCT [11], and LPS [14].

Fig. 6 shows the results with the original methods and their improved versions compared. The results demonstrate that our model significantly improves their performance in terms of different evaluation metrics. To be specific, our model can improve all methods to a similar performance level. For some early methods, the

improved results are comparable to the state-of-the-arts, and for the latest state-of-the-art methods, our model can further improve them to higher performance, which shows the strong ability of our GWB model in improving the quality of salient object detection. A qualitative comparison is also shown in Fig. 7. The background regions are strongly suppressed and the improved maps are more uniform, which demonstrates the effectiveness of the proposed model.

5. Conclusion

In this paper, we propose a novel and unified geodesic weighted Bayesian model for saliency optimization. Motivated by the problems of bottom-up methods and general Bayesian framework, we consider spatial relationship in Bayesian framework by reformulating the computation of observation likelihood, so we can attach more importance to regions which are more likely to be parts of salient object. With that, the background noise is strongly suppressed and the salient objects are highlighted uniformly. Our model is unified and can integrate into all existing methods to improve their performance. Experiments on several benchmark datasets demonstrate that our model can significantly improve the quality of saliency detection to similar performance level, even though the original results are not satisfying, which shows the strong ability of our GWB model in improving the quality of salient object detection.

Acknowledgment

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

Supplementary material

Supplementary material associated with this article can be found, in the online version, at [10.1016/j.patrec.2016.02.008](https://doi.org/10.1016/j.patrec.2016.02.008).

References

- [1] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 1597–1604.
- [2] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Susstrunk, SLIC superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern. Anal. Mach. Intell. 34 (2012) 2274–2282.
- [3] C. Aytekin, S. Kiranyaz, M. Gabbouj, Automatic object segmentation by quantum cuts, in: Proceedings of the 22nd International Conference on Pattern Recognition (ICPR), IEEE, 2014, pp. 112–117.
- [4] C. Aytekin, E.C. Ozan, S. Kiranyaz, M. Gabbouj, Visual saliency by extended quantum cuts, in: Proceedings of IEEE International Conference on Image Processing (ICIP), IEEE, 2015, pp. 1692–1696.
- [5] S. Buoncompagni, D. Maio, D. Maltoni, S. Papi, Saliency-based keypoint selection for fast object detection and matching, Pattern Recognit. Lett. (2015) 32–40.
- [6] M.-M. Cheng, G.-X. Zhang, N.J. Mitra, X. Huang, S.-M. Hu, Global contrast based salient region detection, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 409–416.
- [7] K. Fu, I.Y. Gu, Y. Yun, C. Gong, J. Yang, Graph construction for salient object detection in videos, in: Proceedings of the 22nd International Conference on Pattern Recognition (ICPR), IEEE, 2014, pp. 2371–2376.
- [8] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern. Anal. Mach. Intell. 20 (1998) 1254–1259.
- [9] Y. Jia, M. Han, Category-independent object-level saliency detection, in: Proceedings of IEEE International Conference on Computer Vision (ICCV), 2013, pp. 1761–1768.
- [10] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, S. Li, Automatic salient object segmentation based on context and shape prior, in: Proceedings of British Machine Vision Conference (BMVC), vol. 3, 2011, p. 7.
- [11] J. Kim, D. Han, Y.-W. Tai, J. Kim, Salient region detection via high-dimensional color transform, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 883–890.
- [12] A. Kocak, K. Cizmeciler, A. Erdem, E. Erdem, Top down saliency estimation via superpixel-based discriminative dictionaries, in: Proceedings of the British Machine Vision Conference, BMVA Press, 2014.

- [13] P. Krähenbühl, V. Koltun, Efficient inference in fully connected CRFs with gaussian edge potentials, in: Proceedings of Neural Information Processing Systems (NIPS), 2011.
- [14] H. Li, H. Lu, Z. Lin, X. Shen, B. Price, Inner and Inter Label Propagation: Salient Object Detection in the Wild, *IEEE Trans. Image Processing*, 24 (2015) 3176–3186.
- [15] X. Li, H. Lu, L. Zhang, X. Ruan, M.-H. Yang, Saliency detection via dense and sparse reconstruction, in: Proceedings of IEEE International Conference on Computer Vision (ICCV), IEEE, 2013, pp. 2976–2983.
- [16] R. Liu, J. Cao, Z. Lin, S. Shan, Adaptive partial differential equation learning for visual saliency detection, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 3866–3873.
- [17] T. Liu, J. Sun, N.-N. Zheng, X. Tang, H.-Y. Shum, Learning to detect a salient object, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007.
- [18] S. Manen, M. Guillaumin, L.V. Gool, Prime object proposals with randomized Prim's algorithm, in: Proceedings of IEEE International Conference on Computer Vision (ICCV), 2013, pp. 2536–2543.
- [19] T. Ojala, M. Pietikainen, D. Harwood, Performance evaluation of texture measures with classification based on Kullback discrimination of distributions, *Pattern Recognit.* 1 (1994) 582–585.
- [20] T. Ojala, M. Pietikainen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, *Pattern Recognit.* 29 (1996) 51–59.
- [21] F. Perazzi, P. Krahenbuhl, Y. Pritch, A. Hornung, Saliency filters: contrast based filtering for salient region detection, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 733–740.
- [22] R.B. Potts, Some generalized order-disorder transformations, in: *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 48, Cambridge University Press, 1952, pp. 106–109.
- [23] E. Rahtu, J. Kannala, M. Salo, J. Heikkilä, Segmenting salient objects from images and videos, in: Proceedings of European Conference on Computer Vision (ECCV), Springer, 2010, pp. 366–379.
- [24] C. Rother, V. Kolmogorov, A. Blake, Grabcut: interactive foreground extraction using iterated graph cuts, *ACM Trans. Graph.* 23 (2004) 309–314.
- [25] U. Rutishauser, D. Walther, C. Koch, P. Perona, Is bottom-up attention useful for object recognition? in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, 2004.
- [26] X. Wang, H. Ma, X. Chen, Geodesic weighted Bayesian model for salient object detection, in: Proceedings of IEEE International Conference on Image Processing (ICIP), 2015, pp. 397–401.
- [27] Y. Wei, F. Wen, W. Zhu, J. Sun, Geodesic saliency using background priors, in: Proceedings of European Conference on Computer Vision (ECCV), 2012, pp. 29–42.
- [28] Y. Xie, H. Lu, M.-H. Yang, Bayesian saliency via low and mid level cues, *IEEE Trans. Image Process.* 22 (2013) 1689–1698.
- [29] Q. Yan, L. Xu, J. Shi, J. Jia, Hierarchical saliency detection, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 1155–1162.
- [30] C. Yang, L. Zhang, H. Lu, X. Ruan, M.-H. Yang, Saliency detection via graph-based manifold ranking, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2013, pp. 3166–3173.
- [31] J. Yang, M.-H. Yang, Top-down visual saliency via joint CRF and dictionary learning, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2012, pp. 2296–2303.
- [32] Y. Zhai, M. Shah, Visual attention detection in video sequences using spatiotemporal cues, in: Proceedings of ACM International Conference on Multimedia, 2006, pp. 815–824.
- [33] W. Zhu, S. Liang, Y. Wei, J. Sun, Saliency optimization from robust background detection, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 2814–2821.

Supplementary Material: Geodesic Weighted Bayesian Model for Saliency Optimization

Xiang Wang Huimin Ma Xiaozhi Chen

Tsinghua National Laboratory for Information Science and Technology (TNList)
Department of Electronic Engineering, Tsinghua University

In this supplementary material, we present a more detailed comparison between GWB model and the previous conference version GWB-pre. We utilize saliency maps of several methods as prior distribution, and compare the improved results via GWB and GWB-pre: LC (Zhai and Shah (2006)), FT (Achanta et al. (2009)), HC (Cheng et al. (2011)), RC (Cheng et al. (2011)), SF (Perazzi et al. (2012)), GS (Wei et al. (2012)), HDCT (Kim et al. (2014)), and LPS (Li et al. (2015)).

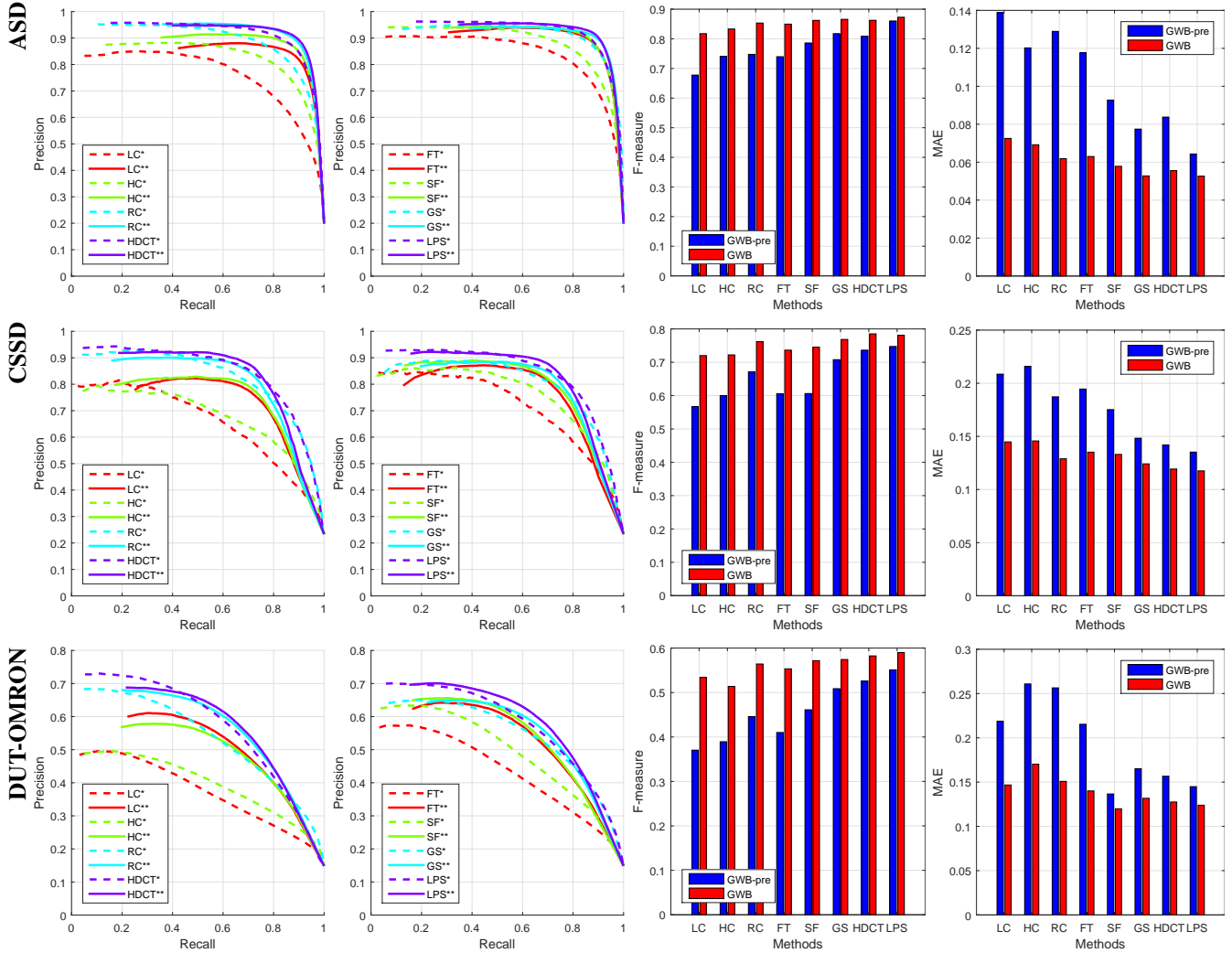


Fig. 1. Comparison between GWB (**) and the previous version GWB-pre(*). The first row are tested on ASD (Achanta et al. (2009)), the second are on CSSD (Yan et al. (2013)), and the third are on DUT-OMRON (Yang et al. (2013)). The first two columns show the comparison of PR curves, the third column shows the comparison of F-measure, and the last column shows the comparison of mean absolute error (MAE). Best viewed in color.

Fig. 1 shows the comparison of different evaluation metrics. We only show the results of GWB and GWB-pre, the results of original methods which are shown in the paper are omitted here for clarity. We can see that our proposed GWB model significantly improved the previous version GWB-pre. In particular, the improved results via GWB have little dependency on the quality of original methods, namely, even though the original maps are not satisfying, the improved maps are also comparable to state-of-the-art, while the previous version GWB-pre has much dependency on the quality of original methods.

References

- Achanta, R., Hemami, S., Estrada, F., Susstrunk, S., 2009. Frequency-tuned salient region detection, in: CVPR, pp. 1597–1604.
- Cheng, M.M., Zhang, G.X., Mitra, N.J., Huang, X., Hu, S.M., 2011. Global contrast based salient region detection, in: CVPR, pp. 409–416.
- Kim, J., Han, D., Tai, Y.W., Kim, J., 2014. Salient region detection via high-dimensional color transform, in: CVPR.
- Li, H., Lu, H., Lin, Z., Shen, X., Price, B., 2015. Inner and inter label propagation: Salient object detection in the wild. arXiv preprint arXiv:1505.07192 .
- Perazzi, F., Krahenbuhl, P., Pritch, Y., Hornung, A., 2012. Saliency filters: Contrast based filtering for salient region detection, in: CVPR, pp. 733–740.
- Wei, Y., Wen, F., Zhu, W., Sun, J., 2012. Geodesic saliency using background priors, in: ECCV, pp. 29–42.
- Yan, Q., Xu, L., Shi, J., Jia, J., 2013. Hierarchical saliency detection, in: CVPR, pp. 1155–1162.
- Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M.H., 2013. Saliency detection via graph-based manifold ranking, in: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE. pp. 3166–3173.
- Zhai, Y., Shah, M., 2006. Visual attention detection in video sequences using spatiotemporal cues, in: ACM Multimedia, pp. 815–824.