J. Vis. Commun. Image R. 46 (2017) 199–207

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

A global and local consistent ranking model for image saliency computation ☆

Yun Xiao [a], Liangmin Wang [b], Bo Jiang [a], Zhengzheng Tu [a], Jin Tang [a,*]

[a] Anhui University, Hefei, China
[b] Jiangsu University, Zhenjiang, China

ARTICLE INFO

Article history:
Received 27 October 2016
Revised 12 March 2017
Accepted 8 April 2017
Available online 9 April 2017

Keywords:
Saliency detection
Markov chain
Graph ranking

ABSTRACT

Image saliency detection is an important issue in computer vision and has been widely used in many applications. In this paper, we propose a new global and local consistent ranking (GLR) model for image saliency computation. Firstly, we propose to use an absorbed Markov chain model to obtain a kind of global ranking for image superpixels, in which the absorbing nodes represent the virtual boundary superpixels and the transient nodes denote the general superpixels of image. Then, the absorbed time from each transient node to boundary absorbing nodes are computed. This absorbing time of transient node measures its global similarity with all absorbing nodes and thus provides a kind of global ranking for each transient node w.r.t. absorbing nodes. At last, we further exploit the local manifold structure and incorporate the local manifold smooth information into ranking process and thus propose a general global and local consistent ranking for saliency detection. Experimental results on several large benchmark databases show the effectiveness of the proposed GLR method.

1. Introduction

Saliency detection, which aims to find the most important and interesting part in an image, is a fundamental problem in computer vision and pattern recognition areas [1–9]. It has been widely applied in many vision tasks including image segmentation [10], object detection and recognition [11], video summarization [12], image retrieval [13], etc.

Many methods have been proposed for saliency detection and computation in computer vision [1,14–19,3,20–22,6,4,7,23,24,9,25]. These methods can be roughly categorized into bottom-up methods, top-down methods [14–21] and mixed models [26, 27,22]. Comparing with top-down methods, we focus on the bottom-up strategy that is generally more flexible and faster to compute and thus has been widely studied and applied in many applications. For example, inspired by biological visual system, Itti et al. [1] bring a simple visual focus (saliency) computation model for natural images. Hou et al. [28] propose a spectral technique to extract backgrounds of images and then obtain the saliency map based on the extracted background results. Han et al. [29] propose to use sparse coding technique to implement a kind of bottom-up saliency computation. Cheng et al. [3] exploit the pixel-wise color contrast to produce saliency maps and propose a histogram-based contrast (HC) method. Perazzi et al. [30] further consider the color distribution in color contrast and introduce a refinement process to achieve a robust saliency detection. Yan et al. [5] present to compute the saliency on different scales of images and then integrate the saliency values of multi-scales simultaneously.

Additionally, more and more graph based learning method have also been proposed in recent years. Harel et al. [31] apply the graph-based visual saliency model and get the feature extraction by using the linear filtering and some fundamental nonlinear filtering as Itti-model, and finally get the saliency value by calculating the contrast between two nodes. Gopalakrishnan et al. [32] construct two kinds of graphs, then extract the global and local properties by using Markov random walk on two graphs and calculate the saliency values of all nodes by contrasting the distance between the node with most saliency value and other one. Jiang et al. [6] propose to use absorbing Markov chain for saliency detection problem in which the saliency value is measured by the absorbing time of transilient nodes in Markov chain. Yang et al. [4] use graph manifold ranking method to learn an optimal ranking function for image superpixels and then measure the saliency value using the ranking scores. Zhu et al. [33] propose boundary connectivity and then utilize a general energy optimization framework to obtain the last saliency map. Li et al. [8] propose a new

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author.
E-mail address: ahu_tj@163.com (J. Tang).

http://dx.doi.org/10.1016/j.jvcir.2017.04.001
1047-3203/

model by using region-based features and image details, and then use a regularized random walk ranking to get final results. Wang et al. [34] exploit a novel approach by exploiting both local graph structure and background priors. Tu et al. [9] present a real-time salient object detection method using the minimum spanning tree.

The recent work has demonstrated that graph based ranking algorithm can usually obtain an effective and desirable saliency detection results for images [4]. However, a limitation for the traditional method is that it generally computes the ranking function based on the local manifold structure of images and thus obviously ignores the global structure of images. Our aim in this paper is to propose a new global and local consistent ranking (GLR) model for image saliency computation. We first utilize to use an absorbed Markov chain model to obtain a kind of ranking for the general superpixels w.r.t. superpixels on image boundary. Then, we compute the absorbed time from each transient node to boundary absorbing nodes. This absorbing time of transient node measures its global similarity with all absorbing nodes, and thus provides a kind of global ranking for each transient node. At last, we further exploit the local manifold structure and incorporate the local manifold smooth information into ranking process and thus propose a general global and local consistent ranking model.

In summary, the contributions of this work mainly include:

(1) We present a new local and global consistent model for saliency detection.
(2) Our model can incorporate both global and local cues simultaneously.
(3) Our model has a closed-form solution.
(4) Experimental results show the effectiveness of our saliency detection model.

2. Global ranking via absorbed time in Markov chain

Recently, ranking methods and algorithms have been widely used for data labeling problem. Formally, let $X = \{x_1, x_2, \ldots, x_l, x_{l+1}, \ldots, x_n\}$ be a dataset containing n data points, where $X_l = \{x_1, x_2, \ldots, x_l\}$ are labeled query data points and the rest $X_u = \{x_{l+1}, x_{l+2}, \ldots, x_n\}$ are unlabeled data. The aim of ranking problem is to determine the ranking function for the test data points X_u according to their relevance to the query data points X_l.

One popular way to address this problem is to use Markov random walks. On this account, a single layer graph $G^g = (V^g, E^g)$ is constructed, where V^g is the node set and E^g denotes the edges between pairs of nodes. The edge weight W_{ij} between node i and j can be calculated by feature vectors of two nodes. Then we get the affinity matrix A [6]. If node i is a transient node, the neighbor node or the neighbor's neighbor node j is connected to node i, $a_{ij} = W_{ij}, a_{ii} = 1$, else $a_{ij} = 0$.

We compute the expected time to absorbtion for each transient node. This expected time can be regarded as the relevances between transient nodes and absorbing nodes. The transient nodes which have similar appearance with absorbing nodes can be absorbed faster, i.e., have less expected time. Therefore, one can use this expected time to determine the order of the transient nodes, i.e., the rest unlabel data points. The expected time is computed as follows [6],

- Step 1. Compute the transition matrix as,

$$P = D^{-1}A, D = \mathrm{diag}\left(\sum_{j=1}^{n} W_{ij}\right) \tag{1}$$

- Step 2. Renumber the nodes so that the first $n-l$ nodes are transient nodes and the last l nodes are absorbing nodes, i.e.,

$$P \to \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix} \tag{2}$$

- Step 3. Compute the expected time as

$$\mathcal{T}_i = \sum_{j=l+1}^{n} [(I-Q)^{-1}]_{ij}. \tag{3}$$

where I is an identity matrix.

We normalize the above $\mathcal{T}_i$ and then calculate the difference with 1 to save as $\mathcal{T} = (\mathcal{T}_{l+1}, \mathcal{T}_{l+2}, \ldots, \mathcal{T}_n)$ that can be regarded as the relevance between the transient node v_i and all the absorbing nodes v_j. Next, we use $\mathcal{T}$ to determine the order/ranking of the unlabel data points X_u w.r.t. the similarities/relevances of the label data points X_l.

3. Global and local consistent ranking

The above Markov chain model provides an effective global ranking for the unlabel data points. However, a main limitation for this method is that its ranking result ignores the intrinsic local manifold structure of data. A natural assumption here is that if two unlabel data points x_i and x_j are close to each other in the intrinsic geometry of the data distribution, then x_i and x_j should also be close in the ranking results. This assumption is also known as manifold learning [35,36]. It motivates us to develop a global and local consistent ranking method by further considering the local geometric relationship between data points.

In order to encode the local structure of data points, a local ranking graph $G^l = (V^l, E^l)$ is first constructed to encode the local neighborhood relationship between data points. Let L_{ij} be the associated weight matrix of graph G^l, then our global and local consistent ranking model (GLR) can be formulated as

$$\min_{r_{l+1}, r_{l+2}, \ldots r_n} J_{GLR} = \sum_{i=l+1}^{n} (r_i - \mathcal{T}_i)^2 + \lambda \sum_{i=l+1}^{n} \sum_{j=l+1}^{n} L_{ij} \left(\frac{r_i}{\sqrt{d_{ii}}} - \frac{r_j}{\sqrt{d_{jj}}} \right)^2, \quad \text{s.t.} \quad r_{l+1}, r_{l+2}, \ldots r_n \geqslant 0 \tag{4}$$

where $\mathcal{T}_i$ is correlation with the normalized value of the absorbed time, and r_i measures the ranking value of node v_i, λ is a parameter which controls the balance of two terms, d_{ii} and d_{jj} are the sum of columns of an affinity matrix obtained by a local ranking graph.

One important benefit of the above GLR model is that it integrates both global and local information in data ranking process. Also, GLR model has a simple closed-form solution, and thus can be computed efficiently. Since $\mathcal{T}_i$ is nonnegative, thus the nonnegative constraint r_i can be ignored. Therefore, the optimal r^* is obtained by solving

$$\frac{\partial J_{GLR}}{\partial r} = 2(r - \mathcal{T}) + 2\lambda \left(r - D^{-\frac{1}{2}} L D^{-\frac{1}{2}} r \right) = 0, \tag{5}$$

where $D = \mathrm{diag}(d_{ii}), d_{ii} = \sum_{j=1}^{n} L_{ij}$.

Thus, we obtain the result,

$$r^* = \left(D - \frac{1}{(1+\lambda)} L \right)^{-1} \mathcal{T} \tag{6}$$

4. Saliency detection with GLR model

In this section, we apply the proposed GLR to saliency detection task. We first use simple linear iterative clustering (SLIC) algorithm [37] to divide the input image into multiple superpixels (the number is $u = n - l$). Each superpixel is viewed as a node in the graph. We use the background prior and object prior to obtain the sal-

iency map. Similar to work [4], we adopt two stages to get the final saliency value of the image. The first stage to get the primary saliency map. By using the first stage result, we binarize the saliency map as foreground prior, then get the final result.

4.1. Graph construction

4.1.1. Global cue graph

For an input image, we construct a single layer graph $G^g = (V^g, E^g)$ with u primary nodes and l duplicated nodes in Fig. 1, where $V^g = \{V_1, V_2, \ldots, V_l, V_{l+1}, V_{l+2}, \ldots, V_n\}$ is the node set and E^g denotes the edges between pairs of nodes. SLIC algorithm divides the input image into u superpixels as primary nodes. Similar to work [6], we duplicated l boundary superpixels as background absorbing nodes, which shows out of the yellow box. The edge E^g is determined as follows [6].

(1) Each transient or absorbing node is connected to the transient nodes which are neighborhood of it or have the same boundaries with its neighboring nodes in the feature space X.
(2) All transient nodes on the boundary are connected to each other.
(3) Any pair of absorbing nodes are unconnected.

The edge weight W_{ij} between node i and j can be calculated as,

$$W_{ij} = e^{-\frac{\|x_i - x_j\|^2}{\sigma^2}}, \tag{7}$$

where x_i and x_j are feature vectors that are extracted from different type features including color value [4], edge and texture information [38]. The color feature includes LAB color values, RGB color values and the hue and the saturation components. The edge feature contains steerable pyramid filters with four directions on three different scales. The texture feature involves Gabor filter responses with 12 orientations and 3 scales. All those 56 features are computed to form a feature vector. σ is a constant.

4.1.2. Local cue graph

The graph generated nodes by SLIC algorithm, which can divide the image into many superpixels. All nodes construct a graph $G^l = (V^l, E^l)$ that is a k-regular graph (we set $k = 2$) [4], where $V^l = \{V_1, V_2, \ldots, V_u\}$. The edge E^l is determined as follows.

(1) Each node connects with not only its neighbors, but also the nodes who have the same neighbors.
(2) We constraint that the boundary nodes are connected to each other.

Then, the weight of edge between node i and j can be calculated by,

$$L_{ij} = e^{-\frac{\|x_i - x_j\|^2}{\sigma^2}}, \tag{8}$$

where x_i and x_j are feature vectors that are extracted from different type features including color value, edge and texture information, and σ is a constant.

4.2. Prior incorporation and two-stage saliency detection

In order to further improve the saliency detection results, we further incorporate some prior information into GLR process. In this section, we incorporate both background and foreground prior information into GLR and propose the two-stage saliency detection process.

4.2.1. Background prior incorporation

Here we use the nodes on the four image boundaries as background priors [4]. Let p^{Bk} $(k = t, b, l, r)$ contain the background priors for the top, bottom, left and right boundaries, respectively. Then, we compute the optimal ranking results r^{Bk} under these priors as following,

$$\begin{aligned} r^{Bk} = \operatorname{argmin} & \sum_{i=l+1}^{n} (r_i - \mathcal{T}_i)^2 + \lambda \sum_{i=l+1}^{n} \sum_{j=l+1}^{n} L_{ij} \left(\frac{r_i}{\sqrt{d_{ii}}} - \frac{r_j}{\sqrt{d_{jj}}} \right)^2 \\ & + \mu \sum_{i=l+1}^{n} (r_i - p_i^{Bk})^2, \\ \text{s.t.} \quad & r_{l+1}, r_{l+2}, \ldots r_n \geqslant 0 \end{aligned} \tag{9}$$

Thus, we obtain the result,

$$r^{Bk} = \left[D - \frac{\mu}{(1 + \lambda + \mu)} L \right]^{-1} (\mathcal{T} + \mu p^{Bk}), \quad k = t, b, l, r, \tag{10}$$

where $D = \operatorname{diag}(d_{ii}), d_{ii} = \sum_{j=1}^{n} L_{ij}$. At last, we combine the four ranking results executing the following process,

$$S_i^B = S_i^{Bt} \times S_i^{Bb} \times S_i^{Bl} \times S_i^{Br}, \quad i = 1, 2, \ldots, n. \tag{11}$$

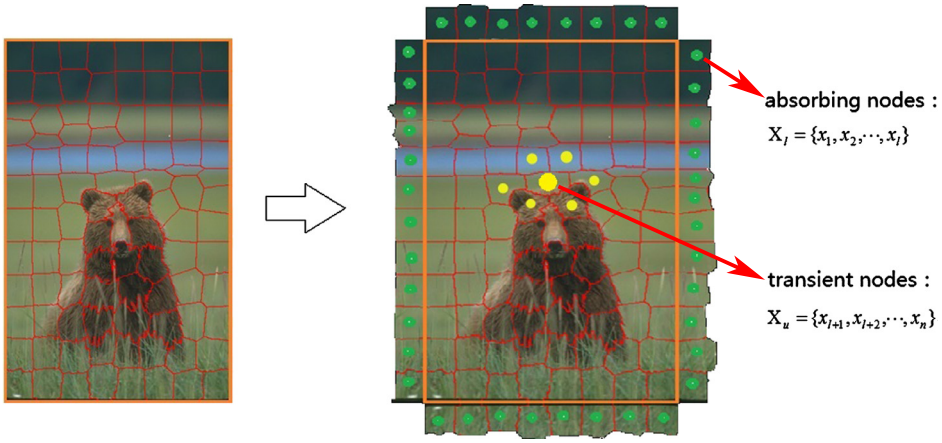


Fig. 1. Left: Superpixels are gained by SLIC algorithm on the original image; Right: The absorbing and transient nodes in the global cue graph. The green nodes are absorbed nodes. The nodes in the yellow box are transient nodes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where

$$S_i^{Bk} = 1 - \frac{r_i^{Bk} - \min\{r^{Bk}\}}{\max\{r^{Bk}\} - \min\{r^{Bk}\}}, \quad k = t, b, l, r. \tag{12}$$

4.2.2. Foreground prior incorporation

As observed in work [4], the above boundary background priors may be incorrect and thus leads to bad saliency results. We further import the foreground prior which is obtained by using the first stage result to overcome this drawback.

As a consequence, we first define the foreground prior $\widetilde{p}^F$ as

$$\widetilde{p}_i^F = \begin{cases} 1, & \text{if } S_i^{Bk} \geqslant mean(S_i^{Bk}) \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

then, we compute the refinement saliency map by solving the following,

$$r^F = \text{argmin} \sum_{i=l+1}^{n} (r_i - \widetilde{\mathcal{T}}_i)^2 + \lambda \sum_{i=l+1}^{n} \sum_{j=l+1}^{n} L_{ij} \left(\frac{r_i}{\sqrt{d_{ii}}} - \frac{r_j}{\sqrt{d_{jj}}} \right)^2 + \mu \sum_{i=l+1}^{n} (r_i - \widetilde{p}_i^F)^2,$$
$$\text{s.t.} \quad r_{l+1}, r_{l+2}, \ldots r_n \geqslant 0 \tag{14}$$

where $\widetilde{\mathcal{T}}_i$ is computed as $\widetilde{\mathcal{T}}_i = 1 - \mathcal{T}_i$. At last, we obtain the final saliency value as the following,

$$S^F = r^F = \left[D - \frac{\mu}{(1+\lambda+\mu)} L \right]^{-1} (\widetilde{\mathcal{T}} + \mu \widetilde{p}^F), \tag{15}$$

5. Experimental results

In this section, we evaluate the effectiveness of our GLR method on several benchmark datasets: DUT-OMRON [4], SOD [39] and SED2 [40]. We compare our method with some recent state-of-art saliency detection methods including MST [9], RR [8], MS [23], COV [46] and SS [41,47], MR [4], LMLC [42], SF [30], CA [43], BM [44], FT [2] and SR [28].

5.1. Implementation details

5.1.1. Parameter setting

The five main parameters of our method are the number of superpixels u, the edge weight control parameter σ, the balance parameters λ and μ. We set $u = 200, \sigma^2 = 0.1$ in our method, λ and μ are separately set to 12.5 and 0.1 in all datasets empirically to obtain better results. Empirically, the proposed GLR is insensitive to these parameters.

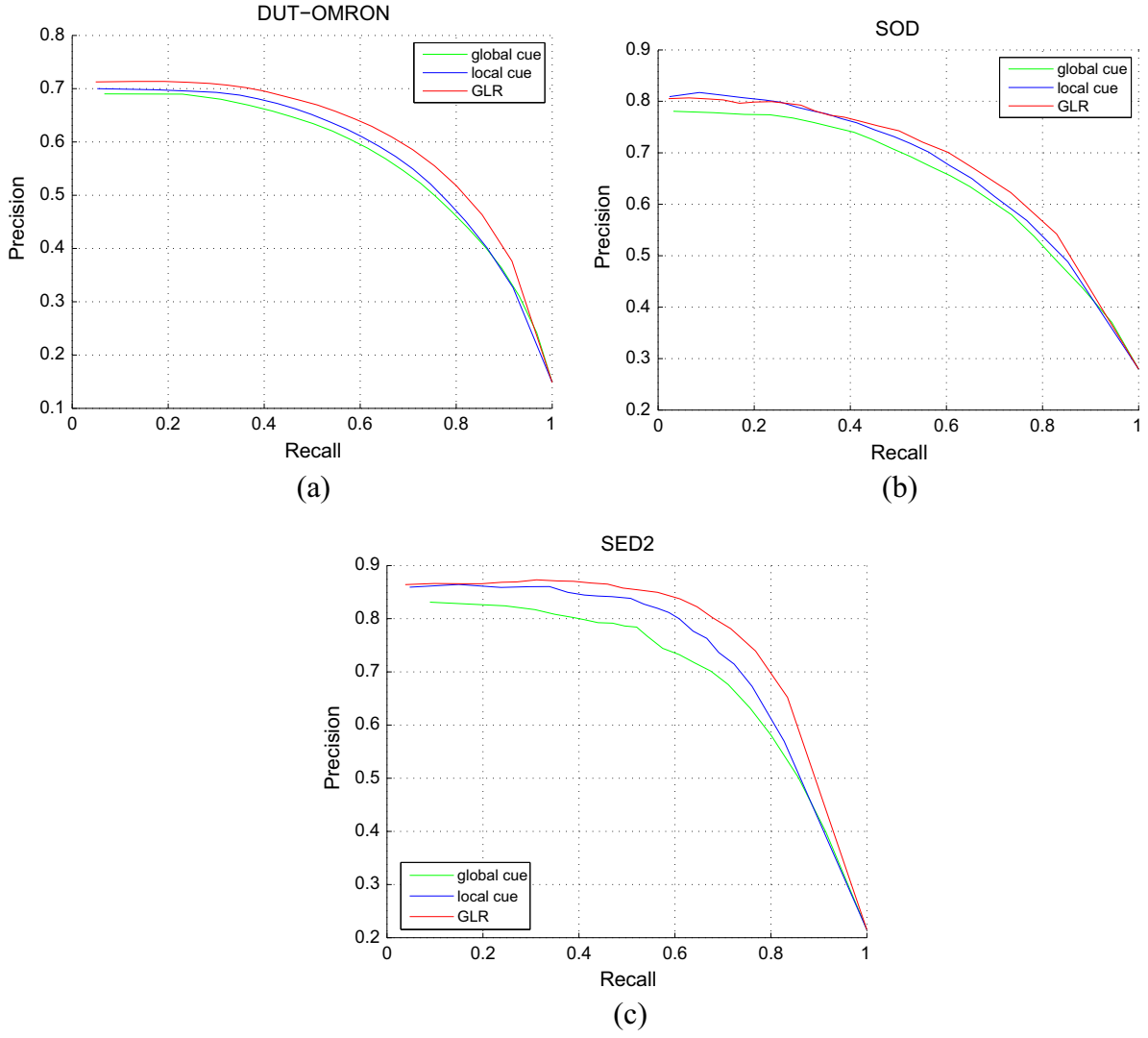


Fig. 2. Precision-recall curves comparing on global and local cue ranking model. (a) DUT-OMRON dataset; (b) SOD dataset; (c) SED2 dataset.

5.1.2. Evaluation metrics

(1) Precision-recall curves: We use the standard precision-recall curves and evaluate the performance of our algorithm. The precision is defined as the ratio of real saliency pixels assigned to all predicted saliency pixels. Recall is calculated as the ratio of the total saliency captured pixels to the ground-truth number. A curve is obtained by the threshold sliding from 0 to 255 to get the comparing between the saliency map and ground truth [4,6].

(2) F-measure: High recall can be achieved at the expense of reducing the precision, so it is important to evaluate both measures simultaneously. The F-measure is computed as the weighted average between the precision and recall values and can be regarded as overall performance measurement. F-measure is defined as,

$$F = \frac{(1+\alpha)Precision \cdot Recall}{\alpha Precision + Recall}, \tag{16}$$

Similar to previous works [2,4,8], we set $\alpha = 0.3$ in our experiments.

(3) Mean absolute error: In additional to F-measure, we also use the mean absolute error(MAE) to calculate the average difference between the saliency map and the ground truth. Both of them are normalized in the range $[0, 1]$. This evaluation method is also

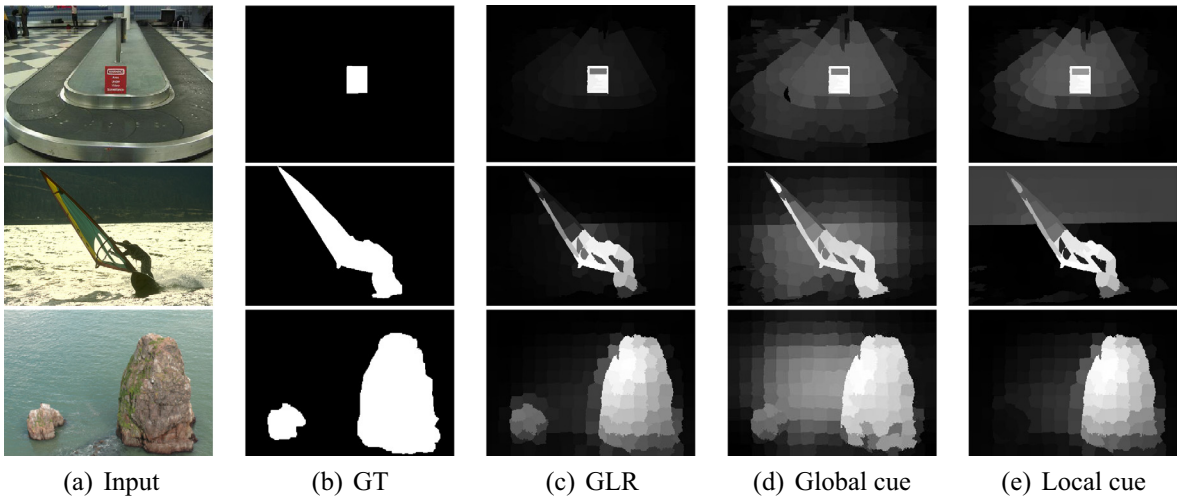


Fig. 3. Some visual comparison examples selected from DUT-OMRON, SOD, SED2 datasets. (a) Input images; (b) ground truth; (c) global and local consistent ranking model results; (d) only global cue ranking results; (e) only local cue ranking results.

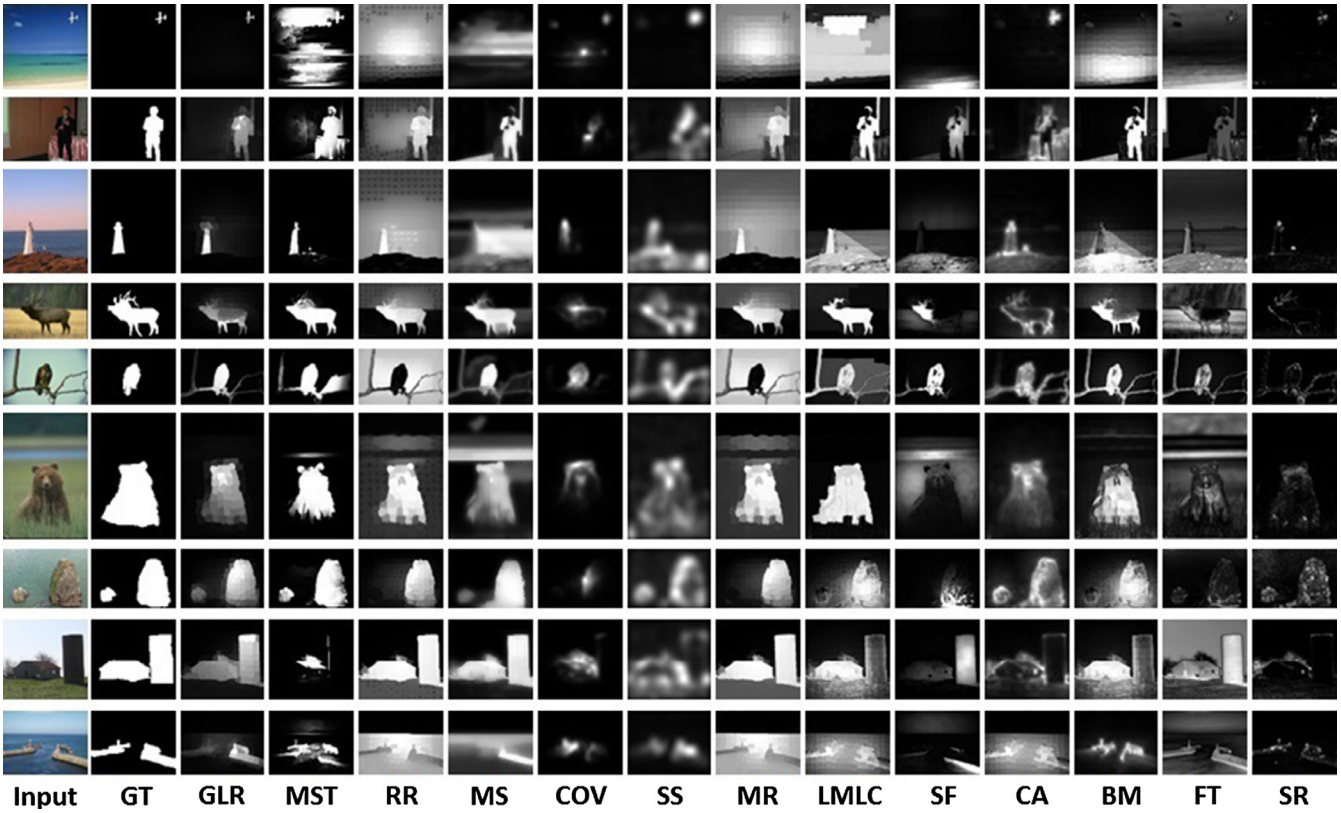


Fig. 4. Examples of output saliency maps results using different algorithms on the DUT-OMRON, SOD and SED2 datasets.

used in many recent methods [41,33,45,9] and plays an important role in different applications such as image segmentation and cropping. Then, compute MAE score as,

$$MAE = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W}|S(i,j) - GT(i,j)|, \quad (17)$$

where H, W denote the height and width of image, respectively. $S(i,j)$ is the saliency value of pixel level, $GT(i,j)$ is the ground truth of pixel level.

5.2. Global and local cue validation

We have experiments on the global and local cue ranking separately on DUT-OMRON, SOD and SED2 datasets. We compute the precision-recall curves on three datasets. Fig. 2 show that the GLR ranking results are better than single global or local cue ranking. Fig. 3 give an example image ranking result by different methods.

5.3. Comparison with state-of-the-art

Fig. 4 shows some visual comparison examples selected from three datasets. The first three rows are obtained from DUT-OMRON dataset. The following three rows are picked out from SOD dataset. The last three rows are cheesed from SED2 dataset. Intuitively, we note that our method can get a most saliency and better result than other methods.

Figs. 5–7 show the comparison results on DUT-OMRON, SOD and SED2, respectively. In each figure, the left side shows the precision-recall curves and the right side gives the F-measure results. Fig. 8 summarizes the MAE results on these datasets. From these results, we can note that: (1) The precision-recall curves in Figs. 5–7 demonstrate that the proposed GLR method outperforms all the compared state-of-art methods as a whole. (2) Also, GLR outperforms than MR method which demonstrates the benefit of global structure information in ranking based saliency detection process. (3) From the results of F-measure, one can note that GLR generally obtains higher F-measure than other methods, which indicates the better average performance of GLR method on conducting saliency detection tasks. (4) GLR generally obtains lower MAE on DUT-OMRON, SOD datasets and SED2 except MST. Although GLR results are not the most effectively method in all evaluation metrics, they are comparable well as a whole.

5.4. Additional feature selection experiments

We have experiments on comparing with only the color feature (LAB values) [4] and different types of visual features including color,edge and texture [4,38]. The color feature includes LAB color

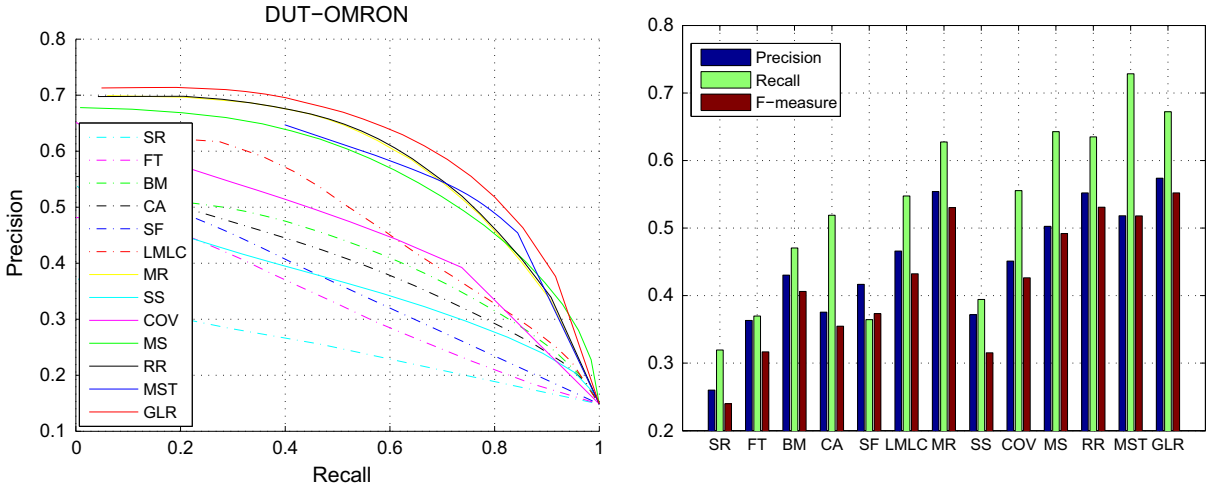


Fig. 5. Left: Precision-recall curves of different methods on DUT-OMRON database. Right: Precision, recall and F-measure histogram statistics.

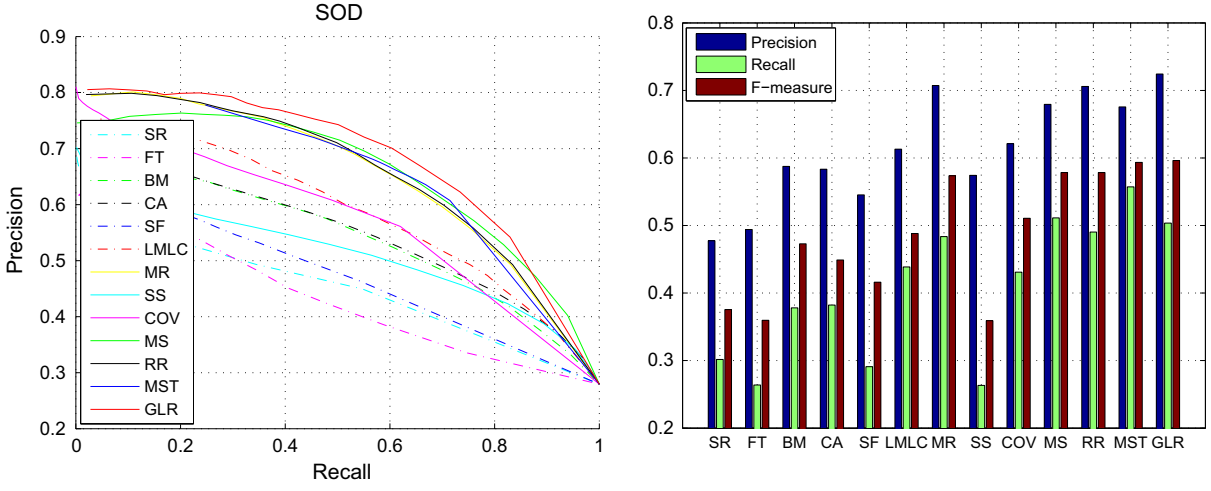


Fig. 6. Left: Precision-recall curves of different methods on SOD database. Right: Precision, recall and F-measure histogram statistics.

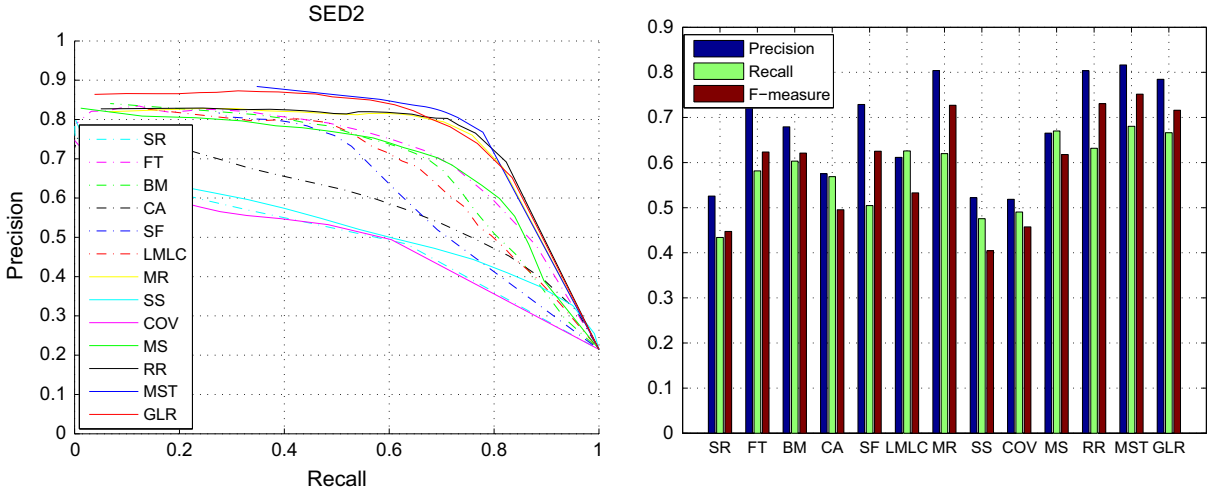


Fig. 7. Left: Precision-recall curves of different methods on SED2 database. Right: Precision, recall and F-measure histogram statistics.

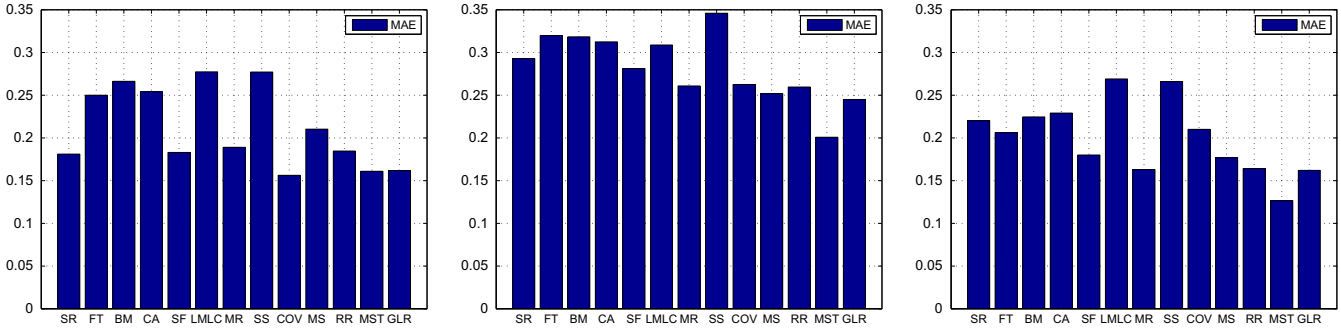
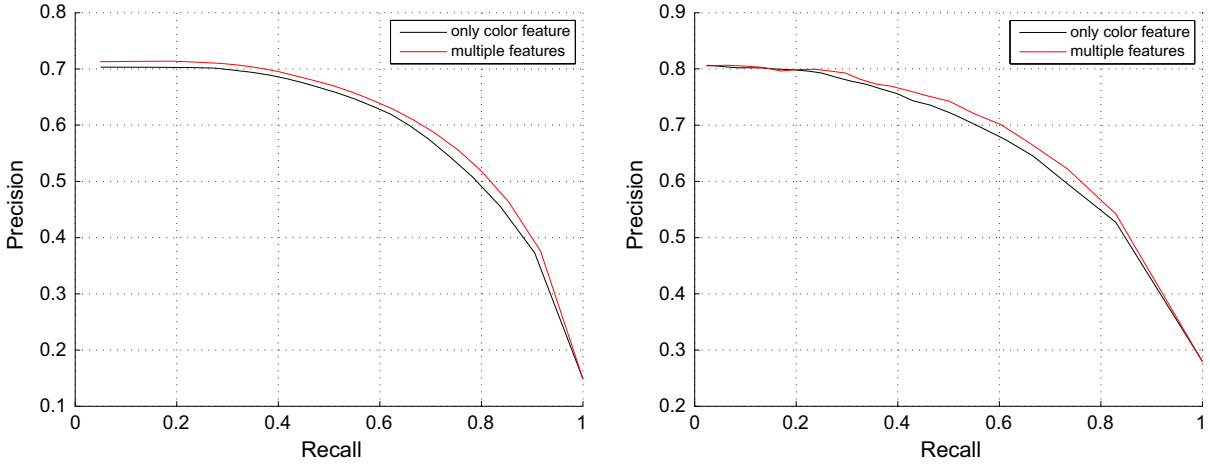


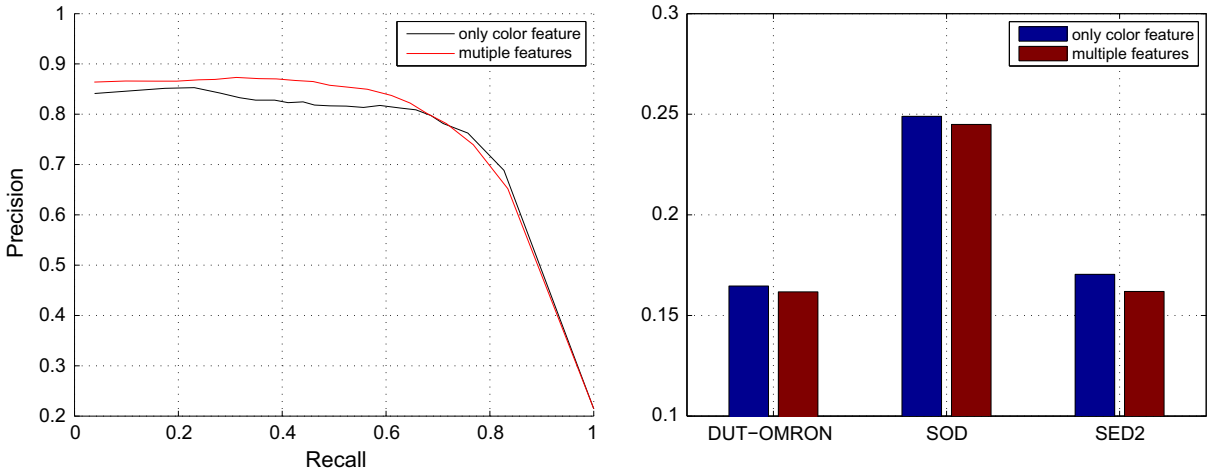
Fig. 8. Comparison of MAE on three datasets. Left: DUT-OMRON database. Middle: SOD database. Right: SED2 database.



(a) PR curves on DUT-OMRON dataset (b) PR curves on SOD dataset

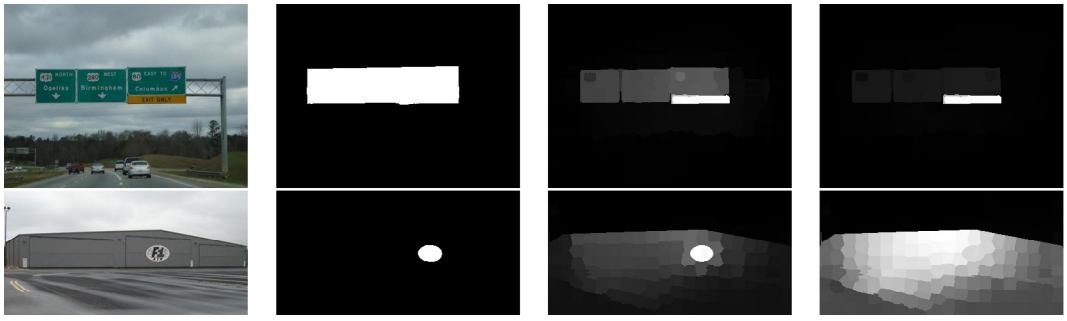
Fig. 9. Precision-recall curves comparing on ranking model by using different features.

values, RGB color values, the hue and the saturation components are extracted for each pixel. Steerable pyramid filters with four directions on three different scales are performed on the image as edge feature [38]. Gabor filter responses with 12 orientations and 3 scales are extracted as texture feature. All those 56 features are computed to form a feature vector. The comparison precision-recall curves and MAE value in Figs. 9 and 10 show that the different types of visual features can get better results. Fig. 11 shows some visual comparison examples selected from DUT-OMRON dataset.



(a) PR curves on SED2 dataset (b) MAE on three different datasets

Fig. 10. Precision-recall curves and MAE comparing on ranking model by using different features.



(a) Input (b) GT (c) Multiple features (d) Color feature

Fig. 11. (a) Input image; (b) the ground truth; (c) the result by using multiple features; (b) the result by using only color feature (LAB value).

6. Conclusion

We have presented a global and local consistent ranking (GLR) algorithm to compute the saliency of images. Comparing with traditional manifold ranking model, GLR integrates both global similarity and local manifold structure of images together and thus compute the saliency maps of images more accurately. The global cue is obtained by calculating the absorbed time from transient nodes to absorbed nodes based on Markov chain model while the local cue is incorporated via a local manifold ranking function. Experimental results show that the proposed method performs better than some recent state-of-the-art saliency detection methods on three benchmark datasets.

In our work, we only use the boundary as background prior. In our future, we will explore some other high-level background prior in our GLR model. Also, we will extend our method to multi-scale form to obtain more consistent saliency results.

Acknowledgments

This work was sponsored by the National Natural Science Foundation of China (Nos. 61472002, 61602001, 61602006), the Natural Science Foundation of Anhui Higher Education Institution of China (KJ2015A110) and by Co-Innovation Center for Information Supply & Assurance Technology, Anhui University.

References

[1] L. Itti, C. Koch, E. Niebur, et al., A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (1998) 1254–1259.
[2] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, IEEE, 2009, pp. 1597–1604.
[3] M.-M. Cheng, G.-X. Zhang, N.J. Mitra, X. Huang, S.-M. Hu, Global contrast based salient region detection, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, IEEE, 2011, pp. 409–416.
[4] C. Yang, L. Zhang, H. Lu, X. Ruan, M.-H. Yang, Saliency detection via graph-based manifold ranking, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, 2013, pp. 3166–3173.
[5] Q. Yan, L. Xu, J. Shi, J. Jia, Hierarchical saliency detection, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, 2013, pp. 1155–1162.
[6] B. Jiang, L. Zhang, H. Lu, C. Yang, M.-H. Yang, Saliency detection via absorbing markov chain, in: International Conference on Computer Vision IEEE International Conference on Computer Vision, 2013, pp. 1665–1672.
[7] X. Li, H. Lu, L. Zhang, X. Ruan, M.-H. Yang, Saliency detection via dense and sparse reconstruction, in: International Conference on Computer Vision IEEE International Conference on Computer Vision, 2013, pp. 2976–2983.
[8] C. Li, Y. Yuan, W. Cai, Y. Xia, D.D. Feng, Robust saliency detection via regularized random walks ranking, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, 2015, pp. 2710–2717.
[9] W.-C. Tu, S. He, Q. Yang, S.-Y. Chien, Real-time salient object detection with a minimum spanning tree, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, 2016, pp. 2334–2342.
[10] L. Wang, J. Xue, N. Zheng, G. Hua, Automatic salient object extraction with contextual cue, in: International Conference on Computer Vision IEEE International Conference on Computer Vision, IEEE, 2011, pp. 105–112.
[11] V. Navalpakkam, L. Itti, An integrated model of top-down and bottom-up attention for optimizing detection speed, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2006, vol. 2, IEEE, 2006, pp. 2049–2056.

[12] Y.-F. Ma, L. Lu, H.-J. Zhang, M. Li, A user attention model for video summarization, in: Proceedings of the Tenth ACM International Conference on Multimedia, ACM, 2002, pp. 533–542.
[13] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, S.-M. Hu, Sketch2Photo: internet image montage, ACM Trans. Graph. (TOG) 28 (5) (2009) 124.
[14] A.Y. Ng, M.I. Jordan, Y. Weiss, On spectral clustering: analysis and an algorithm, Proc. Adv. Neural Inform. Process. Syst. 14 (2002) 849–856.
[15] D. Gao, N. Vasconcelos, Discriminant saliency for visual recognition from cluttered scenes, Adv. Neural Inform. Process. Syst. 17 (2004) 481–488.
[16] S. Frintrop, G. Backer, E. Rome, Goal-directed search with a top-down modulated computational attention system, in: Dagm Conference on Pattern Recognition, 2005, pp. 117–124.
[17] C. Kanan, M.H. Tong, L. Zhang, G.W. Cottrell, Sun: top-down saliency using natural statistics, Vis. Cogn. 17 (6-7) (2009) 979–1003.
[18] L. Marchesotti, C. Cifarelli, G. Csurka, A framework for visual saliency detection with applications to image thumbnailing, in: International Conference on Computer Vision IEEE International Conference on Computer Vision, 2009, pp. 2232–2239.
[19] B. Alexe, T. Deselaers, V. Ferrari, What is an object?, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, IEEE, 2010, pp 73–80.
[20] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, H.-Y. Shum, Learning to detect a salient object, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2) (2011) 353–367.
[21] J. Yang, M.-H. Yang, Top-down visual saliency via joint CRF and dictionary learning, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, IEEE, 2012, pp. 2296–2303.
[22] A. Borji, D.N. Sihite, L. Itti, Probabilistic learning of task-specific visual attention, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, IEEE, 2012, pp. 470–477.
[23] N. Tong, H. Lu, L. Zhang, X. Ruan, Saliency detection with multi-scale superpixels, Sign. Process. Lett. IEEE 21 (9) (2014) 1035–1039.
[24] L. Xu, L. Zeng, H. Duan, An effective vector model for global-contrast-based saliency detection, J. Vis. Commun. Image Represent. 30 (2015) 64–74.
[25] Z. Chen, H. Wang, L. Zhang, Y. Yan, H.-Y.M. Liao, Visual saliency detection based on homology similarity and an experimental evaluation, J. Vis. Commun. Image Represent. 40 (2016) 251–264.
[26] L. Itti, C. Koch, Feature combination strategies for saliency-based visual attention systems, J. Electron. Imag. 10 (1) (2001) 161–169.
[27] V. Navalpakkam, L. Itti, An integrated model of top-down and bottom-up attention for optimizing detection speed, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006, pp. 2049–2056.
[28] X. Hou, L. Zhang, Saliency detection: a spectral residual approach, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007, IEEE, 2007, pp. 1–8.
[29] B. Han, H. Zhu, Y. Ding, Bottom-up saliency based on weighted sparse coding residual, in: International Conference on Multimedea 2011, Scottsdale, AZ, USA, November 28–December, 2011, pp. 1117–1120.
[30] F. Perazzi, P. Krähenbühl, Y. Pritch, A. Hornung, Saliency filters: contrast based filtering for salient region detection, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, IEEE, 2012, pp. 733–740.
[31] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, in: Advances in Neural Information Processing Systems, 2006, pp. 545–552.
[32] V. Gopalakrishnan, Y. Hu, D. Rajan, Random walks on graphs for salient object detection in images, IEEE Trans. Image Process. 19 (12) (2010) 3232–3242.
[33] W. Zhu, S. Liang, Y. Wei, J. Sun, Saliency optimization from robust background detection, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, 2014, pp. 2814–2821.
[34] Q. Wang, W. Zheng, R. Piramuthu, GraB: visual saliency via novel graph model and background priors, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, 2016, pp. 535–543.
[35] M. Belkin, P. Niyogi, V. Sindhwani, Manifold regularization: a geometric framework for learning from labeled and unlabeled examples, J. Mach. Learn. Res. 7 (2006) 2399–2434.
[36] S. Li, L. Li, Y. Fu, Low-Rank and Sparse Dictionary Learning, Springer International Publishing, 2014.
[37] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Ssstrunk, Slic superpixels, Epfl, Technical Report 149300, 2010.
[38] X. Shen, Y. Wu, A unified approach to salient object detection via low rank matrix recovery, in: IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 853–860.
[39] V. Movahedi, J.H. Elder, Design and perceptual validation of performance measures for salient object segmentation, in: Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, IEEE, 2010, pp. 49–56.
[40] S. Alpert, M. Galun, R. Basri, A. Brandt, Image segmentation by probabilistic bottom-up aggregation and cue integration, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007, 2007, pp. 1–8.
[41] A. Borji, M.-M. Cheng, H. Jiang, J. Li, Salient object detection: a benchmark, IEEE Trans. Image Process. 24 (12) (2015) 5706–5722.
[42] Y. Xie, H. Lu, M.-H. Yang, Bayesian saliency via low and mid level cues, IEEE Trans. Image Process. 22 (5) (2013) 1689–1698.
[43] S. Goferman, L. Zelnik-Manor, A. Tal, Context-aware saliency detection, IEEE Trans. Pattern Anal. Mach. Intell. 34 (10) (2012) 1915–1926.
[44] Y. Xie, H. Lu, Visual saliency detection based on bayesian model, Int. Conf. Image Process. 263 (4) (2011) 645–648.
[45] Y. Qin, H. Lu, Y. Xu, H. Wang, Saliency detection via cellular automata, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, 2015, pp. 110–119.
[46] E. Erdem, A. Erdem, Visual saliency estimation by nonlinearly integrating features using region covariances, J. Vision 13 (4) (2013) 1–20.
[47] X. Hou, J. Harel, C. Koch, Image signature: highlighting sparse salient regions, IEEE TPAMI 34 (1) (2012) 194–201.