Robust Saliency Detection via Regularized Random Walks Ranking

Changyang Li[1], Yuchen Yuan[1], Weidong Cai[1], Yong Xia[2], and David Dagan Feng[1]
[1]The University of Sydney [2]Northwestern Polytechnical University

Abstract

In the field of saliency detection, many graph-based algorithms heavily depend on the accuracy of the pre-processed superpixel segmentation, which leads to significant sacrifice of detail information from the input image. In this paper, we propose a novel bottom-up saliency detection approach that takes advantage of both region-based features and image details. To provide more accurate saliency estimations, we first optimize the image boundary selection by the proposed erroneous boundary removal. By taking the image details and region-based estimations into account, we then propose the regularized random walks ranking to formulate pixel-wised saliency maps from the superpixel-based background and foreground saliency estimations. Experiment results on two public datasets indicate the significantly improved accuracy and robustness of the proposed algorithm in comparison with 12 state-of-the-art saliency detection approaches.

1. Introduction

Human saliency is usually referred as local contrast [17], which typically originates from contrasts between an item and its surroundings, such as differences in color, texture, shape, etc. This mechanism measures intrinsically salient stimuli to the vision system that primarily attracts human attention in the early stage of visual exposure to an input image [16]. Intermediate and higher visual processes may automatically judge the importance of different regions of the image, and conduct detailed processes only on the "salient objects" that mostly related to the current tasks, while neglecting the remaining "background" regions [5]. The detection of such salient objects in the image is of significant importance, as it directs the limited computational resources to faster solutions in the subsequent image processing and analysis [6].

In computer vision, salient object detection algorithms can be categorized into bottom-up approaches [1, 6, 8, 11, 13, 17, 20, 27], and top-down approaches [4, 7, 9, 28]. The bottom-up strategy of saliency detection is pre-attentive and data-driven [27]. It is usually fast to execute and easy to adapt to various cases compared to top-down approaches, and therefore has been widely applied. As a pioneer, Itti et al. [17] introduce a conceptually computational model for visual attention detection based on multiple biological feature maps generated by mimicking human visual cortex neurons. Achanta et al. [1] propose a frequency-tuned approach, which results in a saliency map from color differences of the entire image directly. In [6], Cheng et al. present the histogram-based contrast (HC), which exploits the pixel-wise color separation to produce saliency maps, and the region-based contrast (RC), which is an improvement of HC that takes spatial distances into account at the cost of reduced computational efficiency. To overcome the limitations of color contrast, Fu et al. [8] illustrate the workflow of a combined color contrast and color distribution saliency detection algorithm, together with a refinement process to suppress noise and artifacts.

Recently, more graph-based approaches have been proposed. Harel et al. [13] offer the graph based visual saliency (GBVS), a graph-based bottom-up saliency model with dissimilarity measurements to extract saliency information. In the work of Gopalakrishnan et al. [11], the random walks model has been exploited in an automatic salient-region-extraction method to effectively detect the rough location of the most salient object in an image. The method established by Mai et al. [20] formulates saliency by pixel-wise aggregation, conditional random field (CRF) aggregation and image-dependent aggregation. Jiang et al. [19] introduce the discriminative regional feature integration (DRFI), which integrates regional contrast, property and backgroundness descriptor together to formulate the master saliency map. And based on graph-based manifold ranking (MR), the work of Yang et al. [27] utilizes the four boundaries of the input image as background prior to extract foreground queries for the final saliency map.

The results in [27] demonstrate that the MR algorithm outperforms most of the state-of-the-art saliency detection methods and is more computationally efficient. However, there are flaws that hinder it from full performance. Firstly, the four boundaries used as background queries in MR may be implausible for the background saliency detection. In other words, one or more boundaries may be adjacent to the foreground object and undesirable results may

emerge if we still use them as background queries. Another drawback of MR is that it depends on the pre-processed superpixel segmentation, whose inaccuracy may directly lead to the failure of the entire algorithm. Besides, assigning the same saliency value to all pixels in a superpixel node cannot exploit the full potential of the detail information from the original image.

To improve the overall quality of the saliency map, we first filter out one of the four boundaries that most unlikely belong to the background before conducting the background saliency estimation. To improve the foreground saliency estimation, we suggest the proposed regularized random walks ranking, which consists of a pixel-wise graph term [12] and a newly formulated fitting constraint to take local image data and prior estimation into account. This fitting constraint is able to maximally utilize the saliency estimation results from the former steps instead of the selected seed points alone. The regularized random walks ranking is independent of the superpixel segmentation, and can generate pixel-wised saliency maps that reflect full-details of the input images.

The rest of this paper is organized as follows. The basic principles of manifold ranking and random walks are described in Section 2. Section 3 elaborates the three major steps of our proposed approach together with a complete algorithm summary. Section 4 illustrates experiment results on the MSRA10K dataset and the DUT-OMRON dataset. Finally, the conclusion is made in Section 5.

2. Background

In this section, we provide a brief review of the manifold ranking model and the random walks model as preliminary knowledge.

2.1. Manifold Ranking

Manifold ranking is initially used in pattern classification [30, 31]. It assigns ranks to the elements in a dataset that reveal their likelihood being in a certain group with respect to the intrinsic manifold structure. Given a dataset $\chi = \{x_1,...,x_s,x_{s+1},...,x_n\} \in \mathbb{R}^m$, where n is the element number, the first s elements are the labeled queries while the rest are the unknowns, which need to be ranked. This identification is recorded in an indication vector $y = [y_1,...,y_n]^T$, where $y_i = 1$ if x_i belongs to the queries, and $y_i = 0$ otherwise. A graph structure $G = (V,E)$ with nodes V and edges E is then established, where V corresponds to the dataset χ, and E collects all the connections of any two nodes in G quantified by a weight matrix $W = [w_{ij}]_{n\times n}$. The degree matrix $D = diag(d_1,...,d_n)$ is then generated where

$$d_i = \sum\nolimits_j w_{ij}. \tag{1}$$

Let $f : \chi \to \mathbb{R}^n$ be the ranking function assigning rank values $f = [f_1,...,f_n]^T$ to χ, which would be obtained by solving the following minimization problem,

$$f^* = \arg\min_f \frac{1}{2}\left(\sum_{i,j=1}^{n} w_{ij} \left\| \frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_j}} \right\|^2 + \mu \sum_{i=1}^{n} \|f_i - y_i\|^2 \right), \tag{2}$$

where μ is a controlling parameter. The optimized solution is given in [24, 27, 31] as

$$f^* = (D - \alpha W)^{-1} y, \tag{3}$$

where $\alpha = 1/(1+\mu)$.

The manifold ranking model is used to estimate the rough saliency in the proposed method in Section 3. The input image is first segmented into n superpixels via the simple linear iterative clustering (SLIC) approach [3]. A superpixel-based graph $G = (V,E)$ is subsequently constructed with nodes V as superpixels. The edge set E is defined with the following three criteria [27]: 1) Neighboring nodes with shared edges are connected to each other; 2) Each node is also connected to the neighbor nodes of its own neighbors; 3) Any two nodes from the four boundaries of the graph are treated as connected. The weight matrix W is established based on E, in which the weight of adjacent nodes is defined as

$$w_{ij} = \exp\left(-\frac{\|c_i - c_j\|^2}{\sigma^2} \right), \tag{4}$$

where c_i and c_j are the mean CIELab colors of the two nodes i and j, and σ is a controlling constant. The remaining elements of W for the unconnected nodes are all assigned as zeros, and the degree matrix D is computed in Eq.1.

2.2. Random Walks

Random walks is a mathematical formalization of a random sequence path, which leads an element to a seed location with the highest likelihood [12]. Given a dataset $\chi = \{x_1,...,x_n\} \in \mathbb{R}^m$, where n is the element number, the task is to group the elements into K classes. We first mark s elements from χ as the seed nodes with at least one element of each class. Without loss of generality, we assume that the first s elements of χ are the seeds, so that $\chi = [x_M^T, x_U^T]$, in which x_M are the seed nodes

and x_U are the unseeded nodes. The graph $G=(V,E)$, weight matrix W, and degree matrix D are constructed similarly to those in Section 2.1. We further define the $n\times n$ Laplacian matrix L as

$$L_{uv}=\begin{cases} d_u & \text{if } u=v, \\ -w_{uv} & \text{if } x_u \text{ and } x_v \text{ are adjacent nodes,} \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

Note that we use u and v as element subscripts in pixel-wise graphs to differentiate from i and j used in superpixel-wise graphs. Since the edges E are undirected, L is symmetric. Accordingly, we define the label function for seed nodes as

$$Q(x_u)=k,\ k\in\mathbb{Z},\ 0<k\le K. \quad (6)$$

Then let $p^k=\left[p_1^k,\ldots,p_n^k\right]^T$ denote the probability vector of χ for label k, which can similarly be partitioned as $p^k=\left[\left(p_M^k\right)^T,\left(p_U^k\right)^T\right]$. Here p_M^k is for the seed nodes, which has fixed value as

$$p_u^k=\begin{cases} 1 & Q(x_u)=k, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

The optimized p^k is achieved by minimizing the Dirichlet integral [12],

$$\begin{aligned} Dir\left[p^k\right] &= \frac{1}{2}\left(p^k\right)^T L\left(p^k\right) \\ &= \frac{1}{2}\left[\left(p_M^k\right)^T\left(p_U^k\right)^T\right]\begin{bmatrix} L_M & B \\ B^T & L_U \end{bmatrix}\begin{bmatrix} p_M^k \\ p_U^k \end{bmatrix}. \end{aligned} \quad (8)$$

We differentiate $Dir\left[p^k\right]$ with respect to p_U^k, and the critical point is found,

$$p_U^k=-L_U^{-1}B^T p_M^k. \quad (9)$$

In Section 3, the random walks model is reformulated for the final saliency map computation. The graph $G=(V,E)$ is pixel-wise, and the weight matrix W is defined as

$$w_{uv}=\exp\left(-\frac{\left\|g_u-g_v\right\|^2}{\sigma^2}\right), \quad (10)$$

where g_u and g_v are the intensities at pixel u and v, and σ is the same controlling constant used in Eq.4.

3. The Proposed Algorithm

The proposed saliency detection algorithm consists of three major steps. The step one removes the boundary with the lowest probability belonging to the background, and generates saliency estimation via background queries; the step two generates foreground saliency estimation based on the complementary values of the background estimation; the step three extracts seed references from the step two, and calculates the pixel-wise saliency map with the proposed regularized random walks ranking.

3.1. Background Saliency Estimation

As stated in the introduction, it is possible for a boundary in the input image to be occupied by the foreground object. Using such a problematic boundary as queries in the background saliency estimation may lead to undesirable results, and a typical example is illustrated in the second column of Figure 1. We therefore optimize the boundary influences by locating and eliminating erroneous boundaries before the background saliency estimation.

Given the conspicuous difference of color and contrast between the background and the salient object, the erroneous boundary tends to have distinctive color distribution compared to the remaining three. Hence, we treat the superpixel boundaries as connected regions, and calculate their normalized pixel-wise RGB histogram respectively,

$$H_b(h)=\frac{1}{l}\sum_{q=1}^{l}\delta\left(I_q-h\right), \quad (11)$$

where $b\in\{top,bottom,left,right\}$ indicates the four boundary locations; l is the total pixel number in the target region; $h=0,\ldots,255$ is the intensity bin variable; I_q is the intensity value of pixel q; and $\delta(\cdot)$ is the unit impulse function. The red, green and blue channels are calculated separately using 256 bins. We then compute the Euclidean distance of any two of the four histograms,

$$A(b_1,b_2)=\sqrt{\sum_{h=0}^{255}\left[\begin{aligned}&\left(H_{b_1}^{red}(h)-H_{b_2}^{red}(h)\right)^2 \\ &+\left(H_{b_1}^{green}(h)-H_{b_2}^{green}(h)\right)^2 \\ &+\left(H_{b_1}^{blue}(h)-H_{b_2}^{blue}(h)\right)^2\end{aligned}\right]}. \quad (12)$$

The resulted 4×4 matrix A is then summed in column-wise, the maximum of which determines the boundary to be removed. E.g. if the second column sums to be the largest, the bottom boundary will be removed.

The superpixels on each of the three remaining sides of the image will be labeled as ones in the indication vector y in Eq.2, while other nodes as zeros. Three ranking results f_l^* will be achieved afterwards based on Eq.3, where l corresponds to the three remaining locations. Since the ranking results show the background relevance

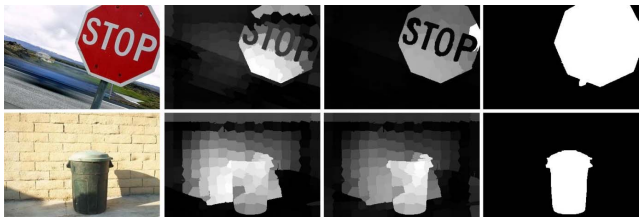


Figure 1. The effect of erroneous boundary removal in Section 3.1. From left to right: input images, background saliency estimations with all boundaries, background saliency estimations after erroneous boundary removal, ground truth.

of each node, we still need to calculate their complement values to obtain the foreground-based saliency,

$$S_l(i) = 1 - f_l^*(i), i = 1,...,n, \tag{13}$$

where n is the total superpixel number. The results are then put into element-by-element multiplication to calculate the saliency result of Section 3.1,

$$S_{step1}(i) = \prod_l S_l(i). \tag{14}$$

The major advantage of erroneous boundary removal is that it helps to relieve the inaccuracy of using all boundaries in cases that one or more of the boundaries happen to be adjacent to the foreground object. As shown in Figure 1, removal of the most irrelevant boundary (right for the first row, and bottom for the second row) leads to more accurate outputs.

3.2. Foreground Saliency Estimation

Section 3.1 calculates the foreground saliency by complementary subtraction of the background saliency estimation, which leads to favorable results in images with conspicuous contrasts between the foreground and the background. However, the background queries alone are sometimes insufficient to fully illustrate the foreground information, especially in cases where the salient object has complicated structure or similar patterns to the background. Subsequent foreground-query-based saliency estimation is hence desired.

The foreground queries are obtained by extracting S_{step1} with a threshold $t = \text{mean}(S_{step1})$, followed by re-performance of Eq.3 with the newly defined indication vector y. The ranking function $\hat{f}$ can be directly calculated from Eq.3 and is treated as the foreground saliency estimation as follows,

$$S_{step2}(i) = \hat{f}(i), i = 1,...,n, \tag{15}$$

which will be used in the next step as seed references.

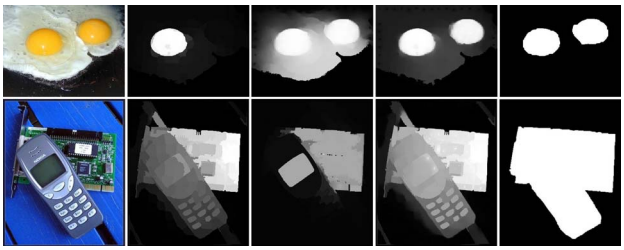


Figure 2. Examples that Eq.16 leads to more precise saliency outputs. From left to right: input images, saliency estimation results, saliency outputs with random walks, saliency outputs with regularized random walks ranking, ground truth.

3.3. Saliency Map Formulation by Regularized Random Walks Ranking

Former manifold-ranking-based saliency detection [27] completely depends on the SLIC superpixel segmentation, which may generate undesirable results if the superpixel segmentation itself is imprecise. In addition, assigning the same saliency value to all pixels within a same node enormously sacrifices the detail information. To overcome these disadvantages, we develop a regularized random walks ranking model to formulate saliency maps, which is independent of the superpixel segmentation, and may reveal pixel-wised saliency map of the input image.

The regularized random walks ranking is extended from the random walks model introduced in Section 2.2. We suggest a fitting constraint, which restricts the Dirichlet integral to be as close to the prior saliency distribution as possible,

$$Dir\left[p^k\right] = \frac{1}{2}\left(p^k\right)^T L\left(p^k\right) + \frac{\mu}{2}(p^k - Y)^T(p^k - Y), \tag{16}$$

where μ is the same controlling parameter used in Eq.2, and Y is a pixel-wise indication vector inheriting the values of S_{step2}. In other words, different pixels within a same superpixel in S_{step2} share the same saliency value in Y. Note that the regularized random walks ranking is computed in pixel-wise, thus both p^k and Y are $N \times 1$ vectors, and L is an $N \times N$ matrix, where N is the total pixel number in the image. We define two thresholds t_{high} and t_{low} as follows,

$$\begin{aligned} t_{high} &= \frac{\text{mean}(S_{step2}) + \max(S_{step2})}{2}, \\ t_{low} &= \text{mean}(S_{step2}), \end{aligned} \tag{17}$$

which are used to select pixels with $Y_u > t_{high}$ as foreground seeds and $Y_u < t_{low}$ as background seeds. The seeds are then combined into p_M^k, $k = 1,2$, where $k = 1$ corresponds to the background label, and $k = 2$ corres-

ponds to the foreground label. The matrix decomposition of Eq.16 is conducted as follows,

$$Dir\left[p^k\right]=\frac{1}{2}\left[\left(p_M^k\right)^T\left(p_U^k\right)^T\right]\begin{bmatrix}L_M & B\\ B^T & L_U\end{bmatrix}\begin{bmatrix}p_M^k\\ p_U^k\end{bmatrix}+\frac{\mu}{2}\left(\begin{bmatrix}p_M^k\\ p_U^k\end{bmatrix}-\begin{bmatrix}Y_M^k\\ Y_U^k\end{bmatrix}\right)^T\left(\begin{bmatrix}p_M^k\\ p_U^k\end{bmatrix}-\begin{bmatrix}Y_M^k\\ Y_U^k\end{bmatrix}\right). \tag{18}$$

After setting the differentiation of $Dir\left[p^k\right]$ with respect to p_U^k as zero, the optimized solution is obtained,

$$p_U^k=\left(L_U+\mu I\right)^{-1}\left(-B^T p_M^k+\mu Y_U^k\right). \tag{19}$$

p_U^k and p_M^k are then combined to form p^k. We set $k=2$ to select the foreground possibility p^2 and reshape it to a matrix S_{final} with same size of the input image as the final foreground saliency output.

Since the seeds are automatically generated from the result of Section 3.2, unlike classical random walks [12], no user interaction is required. The fitting constraint in Eq.16 provides a prior saliency estimation to all pixels instead of the seed pixels alone, which offers a better guidance in calculating the final saliency map. The effect of the fitting constraint in Eq.16 is shown in Figure 2, where the regularized random walks ranking not only greatly improves the saliency map from the previous saliency estimation step, but also remarkably outperforms random walks.

The main process of our proposed algorithm is summarized in Algorithm 1.

Algorithm 1 Saliency Detection by Regularized Random Walks Ranking

Input: An image and related parameters

1: Establish the graph structure with superpixels as nodes; calculate W and D with Eq.4 and Eq.1.

2: Conduct erroneous boundary removal with Eq.12.

3: Acquire the background saliency estimation S_{step1} with Eq.14.

4: Acquire the foreground saliency estimation S_{step2} with Eq.15.

5: Establish the pixel-wise graph structure and obtain L with Eq.5; then compute the saliency possibilities p^k with Eq.19.

6: Set $k=2$ and reshape p^2 into S_{final} as the final saliency output.

Output: a saliency map with the same size as the input image.

4. Experimental Results

Datasets: Our proposed method is tested on two public datasets, the MSRA10K dataset [6] which contains 10,000 randomly-chosen images from the MSRA dataset [5], and the DUT-OMRON dataset [27] with 5,168 manually selected high quality images. Both datasets come with human-labeled ground truth. In our evaluation, we use all of the images in the datasets.

Experimental Parameters Setup:

For the experimental comparison, we use the same parameter settings in [27], where the superpixel number is set to $n=200$, and the two controlling parameters are set to $\sigma^2=0.1$ and $\mu=0.01$, respectively. There is no particular parameter used in the proposed regularized random walks ranking.

Evaluation Metrics:

Precision, recall and F-measure are taken into account in our evaluation with 12 state-of-the-art approaches. The terms are defined in [15] as,

$$precision=\frac{\sum_{i=1}^{N}G(i)\cdot S_{final}(i)}{\sum_{i=1}^{N}S_{final}(i)},\qquad recall=\frac{\sum_{i=1}^{N}G(i)\cdot S_{final}(i)}{\sum_{i=1}^{N}G(i)},\qquad F_\beta=\frac{(1+\beta^2\,precision\cdot recall)}{\beta^2\,precision+recall}, \tag{20}$$

where $G(i)$ is the corresponding ground truth. In other words, precision is the ratio of retrieved salient pixels to all pixels retrieved, and recall is the ratio of retrieved salient pixels to all salient pixels in the image. Precision and recall are usually displayed together as precision-recall curves, which are constructed by binarizing the saliency map with thresholds from 0 to 255. Since these two terms are in general contradictive to each other, the F-measure is adopted as a weighted average between precision and recall. We set $\beta^2=0.3$ to grant more importance to the precision, as suggested in [2].

4.1. Examination of design options

We first examine the major innovations of our proposed algorithm on the MSRA10K dataset, as shown in Figure 3. The blue and green curves illustrate the final saliency output comparison with and without the erroneous boundary removal. Obviously the erroneous boundary removal promotes the curve of the proposed method to a higher level. After that, we generate the saliency maps right after Section 3.2 without using regularized random walks ranking. As shown by the blue and red curves in Figure 3, the complete algorithm also excels the algorithm without using regularized random walks ranking.

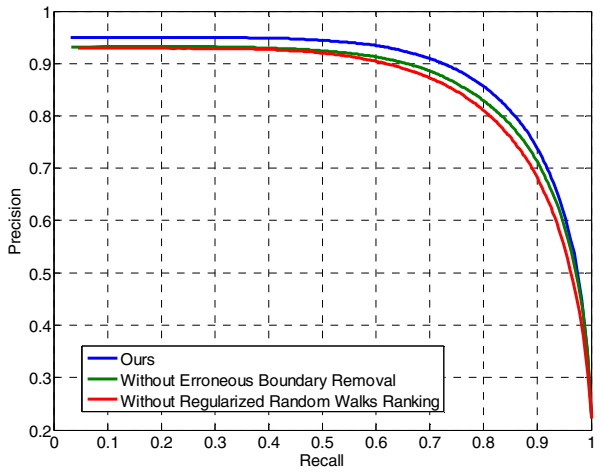


Figure 3. Precision-recall curves on the MSRA10K dataset with different design options of the proposed approach.

Based on the observations above, both the erroneous boundary removal and the regularized random walks ranking have contributions to the overall performance. We therefore adopt both of them in the following evaluations.

4.2. Comparison with State-of-the-art

We then evaluate our proposed algorithm against 12 state-of-the-art saliency detection approaches, namely CA[10], CB[18], FT[2], GS[26], IT[17], LR[25], MR[27], PBO[29], PCA[21], SEG[23], SF[22] and SR[14].

The evaluation is first performed on the MSRA10K dataset, the results of which are shown in Figure 4 and Figure 5. The precision-recall curves in Figure 4(a) and Figure 4(b) demonstrate that the proposed method obviously outperforms all of the state-of-the-art algorithms. The proposed method is especially better than CA and CB, which are two of the top-performance algorithm from a recent benchmark of saliency detection [5]; the proposed method also completely excels its predecessor, i.e. the MR method, which embodies the integrated strength of the improvements we made. On the other hand, Figure 4(c) demonstrates the F-measure comparison; the proposed method achieves the highest F-measure score 0.855, which is 1.06% over the second best algorithm (MR, 0.846). To provide a visual comparison of the different saliency outputs, we choose five of the state-of-the-art methods with the closest performances to the proposed method, namely CB, GS, MR, PCA and LR, and extract example saliency map outputs in Figure 5. We notice that the proposed method generates saliency maps with clearer details and finer boundary adherences.

We further evaluate the proposed algorithm on the DUT-OMRON dataset. The experiment process and evaluation metrics are the same as what we did on the MSRA10K dataset. Precision-recall curves are shown in Figure 6(a) and Figure 6(b), and the F-measure comparison is shown in Figure 6(c). Again, our method outperforms all of the other approaches throughout different precision-recall curves. It also has the optimal F-measure 0.615, which is 0.82% over the second best algorithm (MR, 0.610). Besides the comparison among algorithms, we also notice that the performance of all methods on the DUT-OMRON dataset is in general far poorer than those on the MSRA10K dataset, which indicates that higher performance on more challenging datasets is one of the potential directions of improvement to the proposed method.

4.3. Running Time

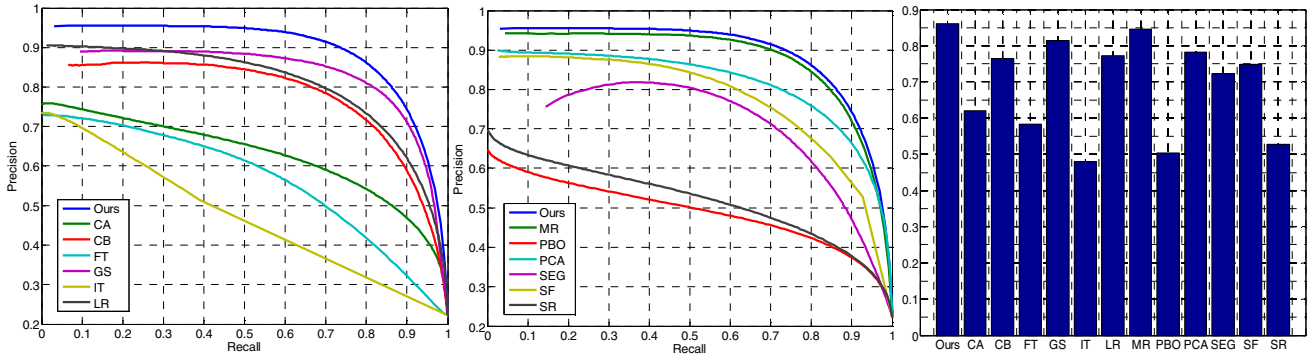
The running time test is conducted on a 64-bit PC with Intel Core i5-4570 CPU @ 3.2 GHz and 8GB RAM. Average running time is computed on the first 1,000 images of the MSRA10K dataset. We choose the five methods with the closest performances to the proposed approach in the test, and the results are shown in Table 1. The proposed algorithm is significantly faster than CB, LR and PCA; and although being slower than MR and GS, our method still outperforms them both considering the overall evaluation performances. The MATLAB implementation of the proposed method is available at our website: `http://sydney.edu.au/engineering/it/~yyua4798/cvpr2015/.`

Method	Ours	CB	GS	MR	PCA	LR
Time(s)	1.12	1.71	0.324	0.869	3.15	13.8

Table 1. Running time test results (seconds per image).

5. Conclusion

In this paper, we propose a novel bottom-up saliency detection method with erroneous boundary removal and regularized random walks ranking. There are two major innovation aspects: firstly, the erroneous boundary removal process effectively eliminates the image boundary with boundary-adjacent foreground superpixels, and thus neutralizes their negative influences in the saliency estimations; secondly, the proposed regularized random walks ranking provides prior saliency estimation to all pixels in the input image, which leads to pixel-wisely detailed and superpixel-independent saliency map outputs. Our approach is fully-automatic without any user supervision requirement. Results of experiments on two public datasets show that the proposed method significantly outperforms 12 state-of-the-art saliency detection algorithms in terms of both accuracy and robustness. In the future, we will further improve the performance of our method, and explore for more potential applications.



(a) (b) (c)

Figure 4. Performance comparison of different algorithms on the MSRA10K dataset. Left and middle: precision-recall curves. Right: average F-measure.

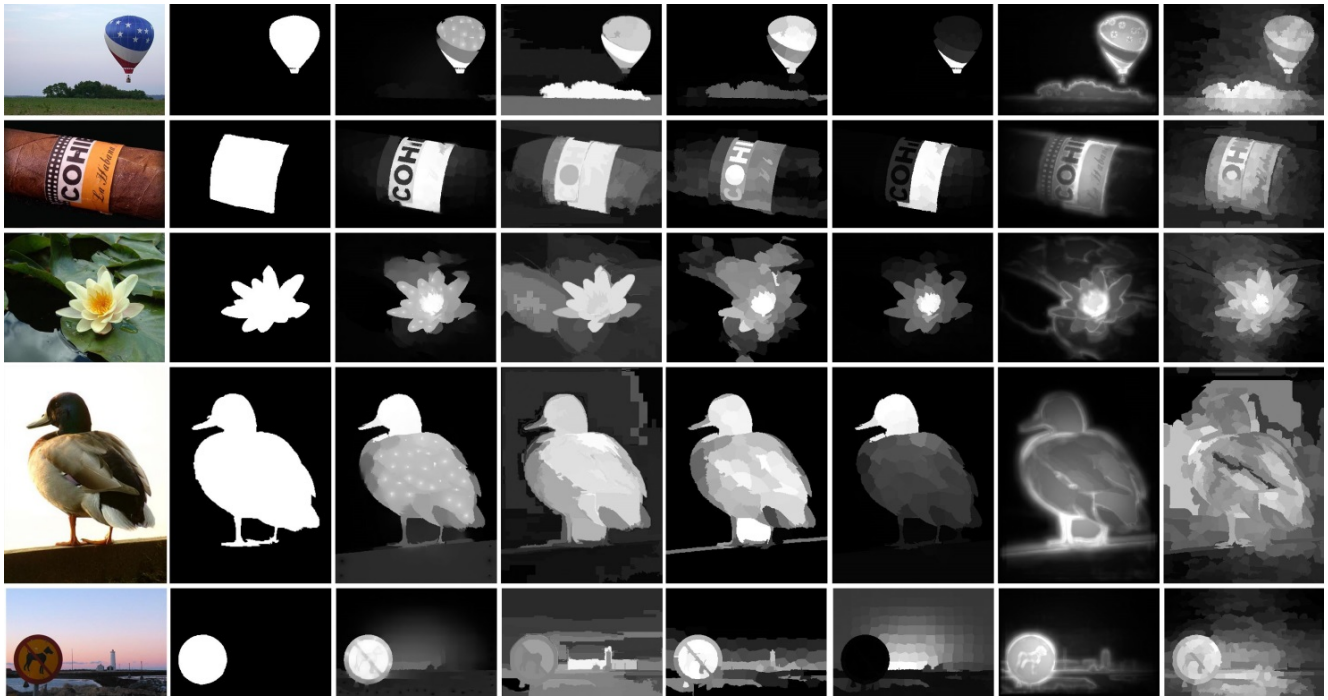
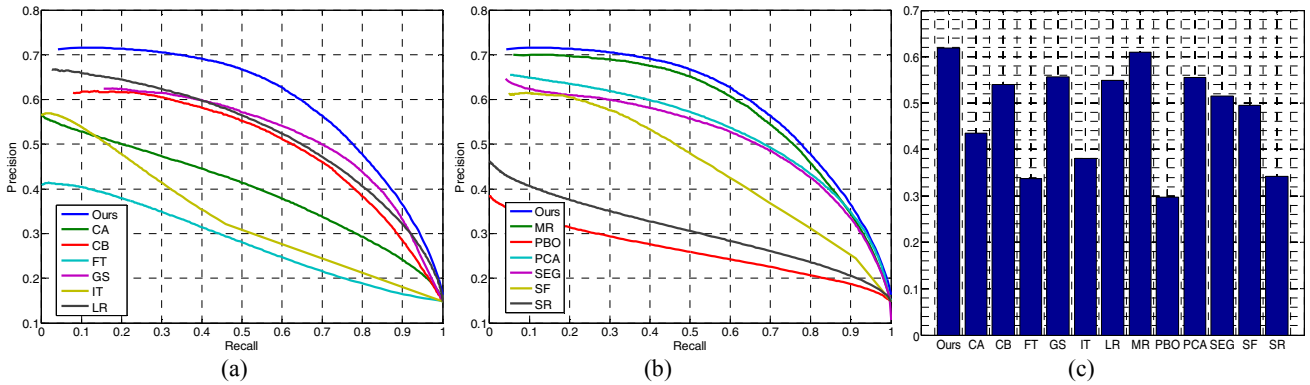


Image Ground truth Ours CB GS MR PCA LR

Figure 5. Examples of output saliency maps using different algorithms on the MSRA10K dataset.



(a) (b) (c)

Figure 6. Performance comparison by using different algorithms on the DUT-OMRON dataset. Left and middle: precision-recall curves. Right: average F-measure.

References

[1] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk, "Salient region detection and segmentation," in *Computer Vision Systems*, ed: Springer, 2008, pp. 66-75.

[2] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 1597-1604.

[3] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 34, pp. 2274-2282, 2012.

[4] A. Borji, D. N. Sihite, and L. Itti, "Probabilistic learning of task-specific visual attention," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 470-477.

[5] A. Borji, D. N. Sihite, and L. Itti, "Salient object detection: A benchmark," in *Computer Vision–ECCV 2012*, ed: Springer, 2012, pp. 414-429.

[6] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011, pp. 409-416.

[7] S. Frintrop, G. Backer, and E. Rome, "Goal-directed search with a top-down modulated computational attention system," in *Pattern Recognition*, ed: Springer, 2005, pp. 117-124.

[8] K. Fu, C. Gong, J. Yang, Y. Zhou, and I. Yu-Hua Gu, "Superpixel based color contrast and color distribution driven salient object detection," *Signal Processing: Image Communication,* vol. 28, pp. 1448-1463, 2013.

[9] D. Gao and N. Vasconcelos, "Discriminant saliency for visual recognition from cluttered scenes," in *Advances in neural information processing systems*, 2004, pp. 481-488.

[10] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 34, pp. 1915-1926, 2012.

[11] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Random walks on graphs for salient object detection in images," *Image Processing, IEEE Transactions on,* vol. 19, pp. 3232-3242, 2010.

[12] L. Grady, "Random walks for image segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 28, pp. 1768-1783, 2006.

[13] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in neural information processing systems*, 2006, pp. 545-552.

[14] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, 2007, pp. 1-8.

[15] G. Hripcsak and A. S. Rothschild, "Agreement, the f-measure, and reliability in information retrieval," *Journal of the American Medical Informatics Association,* vol. 12, pp. 296-298, 2005.

[16] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature reviews neuroscience,* vol. 2, pp. 194-203, 2001.

[17] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 20, pp. 1254-1259, 1998.

[18] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li, "Automatic salient object segmentation based on context and shape prior," in *BMVC*, 2011, p. 7.

[19] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 2013, pp. 2083-2090.

[20] L. Mai, Y. Niu, and F. Liu, "Saliency aggregation: A data-driven approach," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 2013, pp. 1131-1138.

[21] R. Margolin, A. Tal, and L. Zelnik-Manor, "What Makes a Patch Distinct?," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 2013, pp. 1139-1146.

[22] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 733-740.

[23] E. Rahtu, J. Kannala, M. Salo, and J. Heikkilä, "Segmenting salient objects from images and videos," in *Computer Vision–ECCV 2010*, ed: Springer, 2010, pp. 366-379.

[24] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural computation,* vol. 13, pp. 1443-1471, 2001.

[25] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 853-860.

[26] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Computer Vision–ECCV 2012*, ed: Springer, 2012, pp. 29-42.

[27] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 2013, pp. 3166-3173.

[28] J. Yang and M.-H. Yang, "Top-down visual saliency via joint crf and dictionary learning," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 2296-2303.

[29] Y. Zhang, R. Hartley, J. Mashford, and S. Burn, "Superpixels via pseudo-boolean optimization," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, 2011, pp. 1387-1394.

[30] D. Zhou and B. Schölkopf, "Learning from labeled and unlabeled data using random walks," in *Pattern Recognition*, ed: Springer, 2004, pp. 237-244.

[31] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf, "Ranking on data manifolds," *Advances in neural information processing systems,* vol. 16, pp. 169-176, 2004.