Salient Object Detection via Multiple Random Walks

Jiyou Zhai[1,2], Jingbo Zhou[3], Yongfeng Ren[3], Zhijian Wang[1]
[1] College of Computer and Information, Hohai University, Nanjing, 211100, P. R. China
[e-mail: jiyou1018@126.com]
[2] School of Computer Engineering, Nanjing Institute of Technology, Nanjing, 211167, P. R. China
[3] Nanjing University of Information Science & Technology (NUIST), Nanjing, 210044, P. R. China

*Corresponding author: Jiyou Zhai

Abstract

In this paper, we propose a novel saliency detection framework via multiple random walks (MRW) which simulate multiple agents on a graph simultaneously. In the MRW system, two agents, which represent the seeds of background and foreground, traverse the graph according to a transition matrix, and interact with each other to achieve a state of equilibrium. The proposed algorithm is divided into three steps. First, an initial segmentation is performed to partition an input image into homogeneous regions (i.e., superpixels) for saliency computation. Based on the regions of image, we construct a graph that the nodes correspond to the superpixels in the image, and the edges between neighboring nodes represent the similarities of the corresponding superpixels. Second, to generate the seeds of background, we first filter out one of the four boundaries that most unlikely belong to the background. The superpixels on each of the three remaining sides of the image will be labeled as the seeds of background. To generate the seeds of foreground, we utilize the center prior that foreground objects tend to appear near the image center. In last step, the seeds of foreground and background are treated as two different agents in multiple random walkers to complete the process of salient object detection. Experimental results on three benchmark databases demonstrate the proposed method performs well when it against the state-of-the-art methods in terms of accuracy and robustness.

1. Introduction

Saliency detection has attracted intensive attention and achieved considerable progress during the past two decades. Up to now, a great number of detectors based on computational intelligence have been proposed. Although significant progress has been made, it remains a challenging task to develop effective and efficient algorithms for salient object detection.

Saliency models include two main research areas: visual attention which is extensively studied in neuroscience and cognitive modeling, and salient object detection which is of great interest in computer vision. Salient object detection methods can be categorized as bottom-up stimuli-driven [1-19] and top-down task-driven [20-30] approaches. Bottom-up methods are usually based on low-level visual information and are more effective in detecting fine details rather than global shape information. In contrast, top-down saliency models are able to detect objects of certain sizes and categories based on more representative features from training samples. However, the detection results from top-down methods tend to be coarse with fewer details. In terms of computational complexity, bottom-up methods are often more efficient than top-down approaches since they need training process which makes them time consuming.

Saliency models based on bottom-up methods map natural images into saliency maps, in which each image element (e.g., pixel, superpixel and region) is assigned a saliency strength or probability. Early efforts aimed to predict the locations of human eye fixations and introduced the fundamental principles of saliency detection. A representative work by Itti et al. is presented in [4]. They proposed a biologically inspired visual attention model and built a system called neuromorphic vision C++ toolkit. Specifically, they proposed the using of a set of feature maps from three complementary channels as intensity, color, and orientation. The normalized feature maps from each channel were then linearly combined to generate the overall saliency map. Achanta et al. [5] proposed a frequency-tuned approach, which results in a saliency map from color differences of the entire image directly. In [6], Cheng et al. presented 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. [7] illustrated the workflow of a combined color contrast and color distribution saliency detection algorithm, together with a refinement process to suppress noise and artifacts.

Although most previous works mainly concentrated on contrast-based bottom-up saliency, it was believed that at early stage of free viewing, eye movements were mainly directed by bottom-up visual saliency and later on, by high-level factors (e.g., objects [22], actions [23], and events [26]). Thus it was inevitable to combine bottom-up saliency information and top-down factors to build a superior model for predicting eye fixations. Recently, much effort has been made to design discriminative features and saliency priors. Most methods essentially follow the region contrast framework, aiming to design features that better characterize the distinctiveness of an image region with respect to its surrounding area. In [27], three novel features were integrated with a conditional random field. A model based on low-rank matrix recovery was presented in [28] to integrate low-level visual features with higher-level priors. Saliency priors, such as the center prior [27] and the boundary prior [9, 11], were widely used to heuristically combine low-level cues and improve saliency estimation. These saliency priors were either directly combined with other saliency cues as weights [6, 29] or used as

features in learning based algorithms [9, 16, 30]. While these empirical priors can improve saliency results for many images, they may fail when a salient object is off-center or significantly overlaps with the image boundary.

Graph-based approaches, which exploit saliency priors or center priors as saliency seeds or background seeds, have gained much popularity in bottom-up saliency detection and achieved state-of the-art performance. To conduct saliency propagations, an input image is represented by a graph over the segmented superpixels, in which the adjacent superpixels in the image are connected by weighted edges. The saliency values are then iteratively diffused along these edges from the labeled superpixels to their unlabeled neighbors. However, such propagations may incur errors if the unlabeled adjacent superpixels are inhomogeneous or very dissimilar to the labeled ones. For example, [31] formulated the saliency propagation process as random walks on the graph. And based on graph-based manifold ranking (MR), the work of Yang et al. [10] utilized the four boundaries of the input image as background prior to extract foreground queries for the final saliency map. All these methods generate similar propagation sequences which were heavily influenced by the superpixels' spatial relationships. However, once encountering the inhomogeneous or incoherent adjacent superpixels, the propagation sequences were misleading and likely to lead to inaccurate detection results. The results in [10] demonstrated 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. Zhu et al. [11] proposed a novel and reliable background measure, called boundary connectivity, to detect foreground maps in given image. Recently, much more graph-based saliency models which regard saliency detection as the label propagated on a single graph have been proposed [15-17]. All these methods generate similar propagation sequences which are heavily influenced by the super-pixels' spatial relationships. However, once encountering the inhomogeneous or incoherent adjacent superpixels, the propagation sequences are misleading and likely to lead to inaccurate detection results.

In this paper, we propose a novel saliency detection framework via multiple random walkers (MRW) to simulate multiple agents on a graph simultaneously. Those agents traverse the graph according to a transition matrix, but they also interact with one another to achieve a desired goal. The proposed framework is shown in **Fig. 1**. Our algorithm can be segmented into three steps. First, an initial segmentation, i.e., SLIC [32], is required to partition the image into homogeneous regions for measuring saliency. Based on the regions of image, we construct a graph that the superpixel corresponds to the node and the similarity between the neighbors corresponds to the edge. Second, to generate the seeds of background, we first filter out one of the four boundaries that most unlikely belong to the background. The superpixels on each of the three remaining sides of the image will be labeled as the seeds of background. To generate the seeds of foreground, we utilize the center prior that foreground objects tend to appear near the image center. In last step, the seeds of foreground and background are treated as two different agents in multiple random walkers to complete the process of salient object detection. With the random process continues, multiple agents repel the other and form their own dominant regions.

The contributions of this paper are summarized as follows:
(1) We propose a novel salient object detection algorithm which is based on multiple random walkers. Under multiple random walks framework, the seeds of background and foreground are considered as the agents to traverse the graph according to a

transition matrix. As the process goes on, each walker repels the other walker, while forming a dominant region. The region formed by background seeds is considered as the background of the given image.

(2) We design a method of background seeds selection for multiple random walkers in the proposed algorithm. Unlike MR [10] which uses a problematic boundary as the seeds of the background in saliency estimation, we optimize the boundary influences by locating and eliminating erroneous boundaries before the saliency detection.
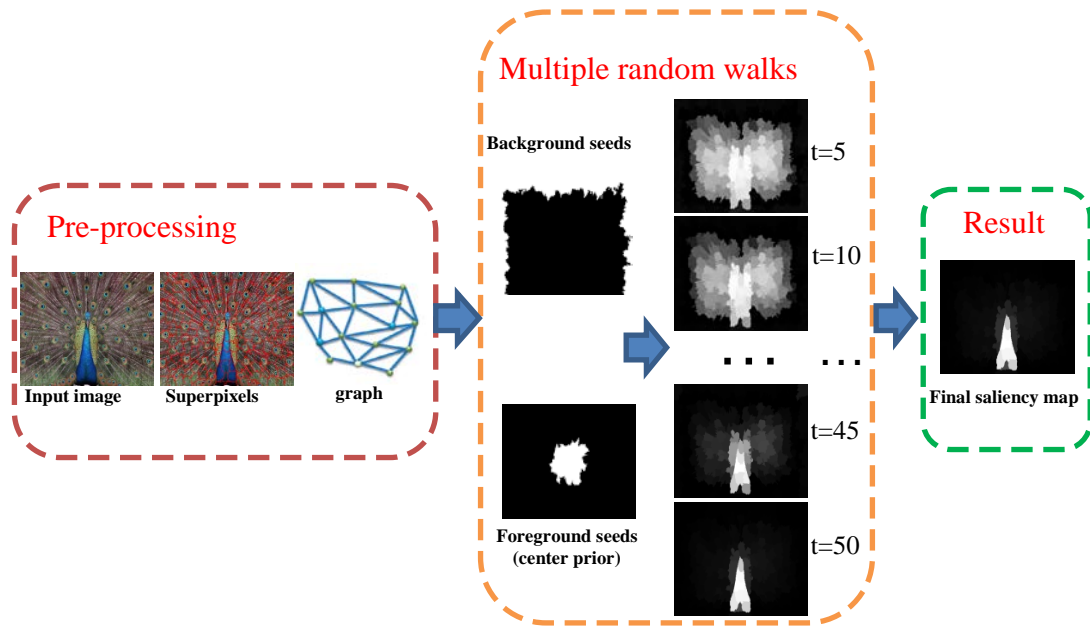

Fig. 1. The pipeline of the proposed algorithm

The remainder of the paper is organized as follows: Section 2 reviews related work which is related to our approach. We demonstrate framework of our saliency detection method in detail in Section 3. Then, we demonstrate our experimental results based on four public image datasets and compare the results with other state-of-art saliency detection methods in Section 4. The final section concludes the paper by summarizing our findings.

2. Related Work

Random walk, a process in which a walker moves randomly from one node to another in a graph, can be used to analyze the underlying data structure of the graph. Previous works on detecting salient regions from images represented as random walk include [12] and [13]. Next, we describe the conventional random walk at first, which is a Markov process of a single random walker.

Let $G = <V, E>$ be an undirected, weighted graph. The node set V consists of data points $x_i, i = 1, 2, \cdots, N$. Edge e_{ij} in the edge set E connects x_i and x_j. Also, let $W \in R^{N \times N}$ be a symmetric matrix, in which the (i, j) th element w_{ij} is the weight of e_{ij}, representing the affinity between x_i and x_j. Typically, w_{ij} is defined as

$$w_{ij} = \begin{cases} \exp\left(-\dfrac{d^2(x_i, x_j)}{\sigma^2}\right) & if \ e_{ij} \in E \\ 0 & otherwise \end{cases} \qquad (1)$$

where d is a dissimilarity function and σ^2 is a scale parameter.

A random walker travels on the graph G. The transition probability a_{ij} that the walker moves from node j to node i is obtained by dividing w_{ij} by the degree of node j, i.e., $a_{ij} = w_{ij} / \sum_k w_{kj}$. In other words, the transition matrix $A = [a_{ij}]$ is computed by normalizing each column of the affinity matrix W. Let $p^{(t)} = [p_1^{(t)}, \cdots, p_N^{(t)}]^T$ be a column vector, in which $p_i^{(t)}$ denotes the probability that the walker is found at node i at time instance t. We then have the temporal recursion

$$p^{(t+1)} = Ap^{(t)} \qquad (2)$$

If the graph has a finite number of nodes and is fully connected, A is irreducible and primitive [33]. Then, the walker has a unique stationary distribution π, satisfying $\pi = A\pi$, and $\pi = \lim_{t \to \infty} p^{(t)}$ regardless of an initial condition $p^{(0)}$. The stationary distribution π conveys useful information about the underlying data structure of the graph, and it is thus exploited in vision applications [14, 34].

Previous works focus on single random walker to detect salient regions from images. In [12], Costa presents two models in which random walks on graphs enable the identification of salient regions by determining the frequency of visits to each node at equilibrium. While some results are presented on only two synthetic images, there is no evaluation of how the method will work on real images. A similar approach in [13] uses edge strengths to represent the dissimilarity between two nodes; the strength between two nodes decreases as the distance between them increases. Here, the most frequently visited node will be most dissimilar in a local context. A major problem when looking for local dissimilarities of features is that cluttered backgrounds will yield higher saliencies as such backgrounds possess high local contrasts. In the work of Gopalakrishnan et al. [14], 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 behavior of random walks on the two separate graphs is used to identify the most salient node of the image along with some background nodes. The final stage of seeded salient region identification uses information about the salient and the background nodes in an accurate extraction of the salient part in the image. To obtain the saliency seeds, they evaluated the 'isolation' of nodes in a global sense and identify nodes corresponding to 'compact' regions in a local sense. Since clutter background often exists in natural images, the dominant seeds located in background were detected as salient seeds, which affected the performance of their method in complex images. Our proposed work is different from them for two aspects. First, our framework is based on the multiple random walks, which take both background and foreground into consideration, diffuse the information of multiple agents on the same graph. Background and foreground can interact with each other until achieve a state of equilibrium. Second, the proposed algorithm exploit a robust method to obtain background seeds which can improve the performance when

process clutter background.

Random walk with restart (RWR) [35] is a generalization of the random walk, in which the walker is compelled to return to specified nodes with a restart probability ε. The RWR recursion is

$$p^{(t+1)} = (1-\varepsilon)Ap^{(t)} + \varepsilon r \tag{3}$$

where $r = [r_1, r_2, \cdots, r_N]^T$ is the restart distribution with $\sum_i r_i = 1$ and $r_i \geq 0$. With probability $1-\varepsilon$, the walker moves ordinarily as in (2). On the other hand, with probability ε, the walker is forced to restart with the distribution r. When $r_i = 1$ and $r_j = 0$ for all $j \neq i$, the stationary distribution of the RWR represents the affinity of each node to the specific node i. This property has been exploited in image segmentation [36] and data mining [35].

Based on the theory of random walk with restart (RWR), Kim et al. [37] proposed a graph-based multiscale saliency detector, which used a coarse scale saliency map to refine a fine scale map. The steady-state distribution obtained in a coarser scale image is used as a restarting vector for the random walk at the higher scale image. In [38], A saliency detection algorithm for video sequences based on the random walk with restart (RWR) adopt RWR to detect spatially and temporally salient regions which use the features of motion distinctiveness, temporal consistency, and abrupt change. The transition probability matrix for the walker uses the spatial features of intensity, color, and compactness. As mentioned before, these two methods are based on single random walker. Our proposed algorithm is also different from them that we average five dissimilarities of node features, including RGB and LAB superpixel means, boundary cues, bag-of-visual-words histograms of RGB and LAB colors to obtain the transition matrix. To achieve saliency detection by information propagation from background and foreground seeds, we employ double random walkers, called foreground walker and background walker to form the saliency map based on the transition matrix.

3. Salient Object Detection by Multiple Random Walks

This section introduces the notion of Multiple Random Walks (MRW). Then we describe how to select the seeds of background and eliminate erroneous boundaries. According to the characteristic of MRW, we define the salient region detection accordingly.

3.1 Multiple Random Walks

The conventional random walk in (2) or (3) describes the movements of a single agent. In contrast, we consider multiple agents who share the same graph and affect the movements of one another.

Suppose there are K agents on a graph. Let $p_k^{(t)}$ denote the probability distribution of agent k at time t. Similar to (3), random movements of agent k are governed by

$$p_k^{(t+1)} = (1-\varepsilon)Ap_k^{(t)} + \varepsilon r_k^{(t)}, k = 1,2,\cdots,K \tag{4}$$

Thus, with probability $1-\varepsilon$, agent k travels on the graph according to the transition matrix A, independently of the other agents. However, with probability ε, agent k visits the nodes

according to the time-varying restart distribution $r_k^{(t)} = [r_{k,1}^{(t)}, \cdots, r_{k,N}^{(t)}]^T$.

We can make the agents interact with one another, by determining the restart distribution as

$$r_k^{(t)} = (1 - \delta^t) r_k^{(t-1)} + \delta^t \phi_k(P^{(t)}) \tag{5}$$

where the function ϕ_k is referred to as the restart rule. It determines a probability distribution ϕ_k from $P^{(t)} = \{p_k^{(t)}\}_{k=1}^K$, which is the set of the probability distributions of all agents at time t.

In (5), δ is a constant within $[0,1]$, called the cooling factor. In an extreme case of $\delta = 0$, the restart distribution becomes time-invariant, and the MRW recursion of each agent in (4) is identical with the RWR recursion in (3). In the other extreme case of $\delta = 1$, $r_k^{(t)} = \phi_k(P^{(t)})$ does not directly depend on the previous distribution $r_k^{(t-1)}$. Suppose that $0 < \delta < 1$. We have

$$\left\| r_k^{(t)} - r_k^{(t-1)} \right\|_\infty = \delta^t \left\| \phi_k(P^{(t)}) - r_k^{(t-1)} \right\|_\infty \le \delta^t \tag{6}$$

Thus, if $s \ge t \ge T$, $\left\| r_k^{(s)} - r_k^{(t)} \right\|_\infty \le \delta^T / (1 - \delta)$. So each element in the restart distribution $r_k^{(t)}$ is a Cauchy sequence in terms of time t. Since a Cauchy sequence in $\mathfrak{R}$ is convergent, the restart distribution $r_k^{(t)}$ converges to a fixed distribution $r_k^{(\infty)}$. Therefore, as t approaches infinity, the MRW recursion in (4) becomes the RWR recursion, and agent k has a stationary distribution eventually. Each agent in the MRW system in (4) and (5) has a stationary distribution [35], given by

$$\pi_k = \lim_{t \to \infty} p_k^{(t)} \tag{7}$$

In our system, multiple walkers adapt their movements based on other walkers' distributions. Then, we extract useful information from the stationary distributions of the multiple walkers.

3.2. Repulsive Restart Rule

By designing the restart rule ϕ_k in (5) for saliency detection, we can simulate two agent (background and foreground, $k = 2$) interactions to achieve a desired goal. For notational simplicity, let us omit the superscripts for time instances. In the MRW system

$$p_k = [p(x_1|\omega_k), \cdots, p(x_N|\omega_k)]^T \tag{8}$$

where $p(x_i|\omega_k)$ is the probability that agent k is found at node i. According to the Bayes' rule, the posterior probability is given by

$$p(\omega_k|x_i) = \frac{p(x_i|\omega_k) p(\omega_k)}{\sum_l p(x_i|\omega_l) p(\omega_l)} \tag{9}$$

which represents the probability that node i is occupied by agent k. The repulsive restart rule

sets the i th element of $\phi_k(\mathrm{P})$ as

$$\phi_{k,i} = \alpha \cdot p(x_i|\omega_k) \cdot p(\omega_k|x_i)) \tag{10}$$

where α is a normalizing factor to make $\phi_k(\mathrm{P})$ a probability distribution. Suppose that foreground agent is dominant at node i, i.e., it has a high posterior probability $p(\omega_k|x_i)$ and a high likelihood $p(x_i|\omega_k)$. Then, it restarts at that node with a high probability, and tends to become more dominant. This has the effect that a foreground agent at a node repels the background agent. The repulsive restart rule in (10) can be rewritten as

$$\phi_k(\mathrm{P}) = \alpha Q_k p_k \tag{11}$$

where Q_k is a diagonal matrix whose (i,i) th element is the posterior probability $p(\omega_k|x_i)$.

For saliency detection, we perform the MRW simulation in (4) and (5), by employing the restart rule in (11), to obtain the stationary distribution π_k of each agent k. Then, node i is assigned a background or foreground label l_i based on the maximum a posteriori (MAP) criterion,

$$l_i = \arg\max_k p(\omega_k|x_i) \tag{12}$$

3.3 Background Seed Selection

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 [11]. 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,

$$H_b(h) = \frac{1}{l}\sum_{q=1}^{n}\delta(I_q - h) \tag{13}$$

where $b \in \{top, bottom, left, right\}$ indicates the four boundary locations; n is the total pixel number in the target region; $h = 0, \cdots, 255$ 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,

$$\varphi(b_1, b_2) = \sqrt{\sum_{h=0}^{255}\left[\left(H_{b_1}^{red}(h) - H_{b_2}^{red}(h)\right)^2 + \left(H_{b_1}^{green}(h) - H_{b_2}^{green}(h)\right)^2 + \left(H_{b_1}^{blue}(h) - H_{b_2}^{blue}(h)\right)^2\right]} \tag{14}$$

The resulted 4×4 matrix φ 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.

3.4 Saliency Detection via MRW

Given an input image, we first construct a graph $G = <V, E>$. The image is initially over-segmented into SLIC superpixels [32], each of which becomes a node in V. For the edge set E, we use the edge connection scheme in [10]. Specifically, each node is connected not only to its spatial neighbors but also to the neighbors of the neighbors, and all boundary nodes at the image border are connected to one another.

For each edge e_{ij}, we determine the affinity weight w_{ij} in (1), by employing the dissimilarity function

$$d(x_i, x_j) = \sum_l \frac{1}{l} d_l(x_i, x_j) \tag{15}$$

We use five dissimilarities d_l of node features, including RGB and LAB superpixel means, boundary cues, bag-of-visual-words histograms of RGB and LAB colors [39]. For the superpixel means of RGB and LAB, we average all feature values in a superpixel node and use the Euclidean distance for measuring the dissimilarity. For the 'Boundary cue,' we use the sum of the gradient magnitudes on the line connecting the centers of two superpixel nodes directly as the dissimilarity. For the 'Bag of-visual-words' histograms of RGB and LAB colors, we build codewords via the k means clustering and construct the codeword histogram for a superpixel node. In this case, we use the Chi-square distance for measuring the dissimilarity. We combine those dissimilarities through averaging them. By normalizing each column of the affinity matrix $W = [w_{ij}]$, we obtain the transition matrix A.

To achieve saliency detection by information propagation from background and foreground seeds, we employ double random walkers, called foreground walker and background walker, whose probability distributions are denoted by p_f and p_b, respectively. These two walkers make interactions according to

$$p_f^{(t+1)} = (1-\varepsilon)A p_f^{(t)} + \varepsilon r_f^{(t)} \tag{16}$$

$$p_b^{(t+1)} = (1-\varepsilon)A p_b^{(t)} + \varepsilon r_b^{(t)} \tag{17}$$

with the repulsive restart rule. It is guaranteed that the probability distributions converge to stationary distributions π_f and π_b, respectively. To sum up, the saliency scores according to saliency and boundary priors can be written as:

$$S(i) = p_f \tag{18}$$

Algorithm 1 describes how to implement the proposed algorithm. The temporal saliency detection consists of three steps. The first step is to segement the image into superpixels and to construct the the graph. In this step, we determine the affinity weight w_{ij} by employing five

dissimilarity functions and then the transition matrix A is determined accordingly. In second step, we compute the seeds of the background and foreground by locating and eliminating erroneous boundaries. In last step, the information of the background and the foreground diffuse in the graph repeatedly until the regions of the foreground stops decreasing.

Algorithm 1. Salient object detection via MRW

Input: Input image I, restart probability ε

1. Segment image I to superpixels

2. Construct a graph $G = <V, E>$, determine the transition matrix A

3. Calculate the seeds of background p_b according to (13)and(14)

4. Detemine the seeds of foreground p_f based on background seeds

5. **repeat**

6. $$p_f^{(t+1)} = (1-\varepsilon)Ap_f^{(t)} + \varepsilon r_f^{(t)}$$

7. $$p_b^{(t+1)} = (1-\varepsilon)Ap_b^{(t)} + \varepsilon r_b^{(t)}$$

8. **until** the regions of foreground stops decreasing

Output: Saliency map $S = p_f$

---

4. Experimental Results and Analysis

In this section, we evaluate the performance of our proposed algorithm (SMRW for short) over several data sets that are widely used in previous works, e.g. [2, 6, 9]. Next, we describe the datasets shortly and report both quantitative and qualitative comparisons of our approach with state-of-the-art approaches in detail. To save the space, we compare our method with several prior ones, including SVO [40], PCA [41], RC [6] and DRFI [9], which are the top four models or their improvements in survey [2]. In addition, we also consider the methods which are based on the three-step procedure, such as DSR [42], LRMR [28], XIE [43] and MR [10], that is not covered in [2] and recently-developed methods, i.e., MS [44], LPS [45], Zhu [11] and BL [46]. In our experiments, we fix the restart probability ε in (16) and (17) to 0.2 for all datasets.

4.1 Datasets and Evaluation Measures

The proposed method is evaluated on three publicly-available datasets with a ground truth in the forms of accurate human-marked labels for the salient regions. (1) MSRA [5] includes 1000 images, originally containing labeled rectangles from nine users drawing a bounding box around what they consider the most salient object. There is a large variation among images including natural scenes, animals, indoor, out-door, etc. We use the salient object (contour) in [9] as binary masks. (2) SED [47] contains two subsets, the first of which is a single-object database (SED1) with 100 color images and only one salient object in each image. The second is a two-object database (SED2), which also has 100 color images with two salient objects in each image. Pixel-wise ground truth annotation for the salient objects in both SED1 and SED2 are provided. (3) SOD [48] is a collection of salient object boundaries based on the Berkeley segmentation data set. Seven subjects are asked to choose the salient object(s) in 300 images. This data set contains many images with multiple objects making it challenging.

We evaluate the performance using the measures used in [2], the PR (precision-recall) curve. Precision is the fraction of detected salient pixels belonging to the salient object in the ground

truth, and recall corresponds to the percentage of salient pixels correctly assigned. The PR curve is created by varying the saliency threshold that determines if a pixel is on the salient object. To obtain F-Measure, we follow [5, 6] to segment a saliency map by the threshold defined as follows:

$$\tau = \frac{2}{H * W} \sum\nolimits_{x=1}^{H} \sum\nolimits_{y=1}^{W} S(x, y) \tag{19}$$

where W and H are the width and height of the saliency map in pixels, respectively, and $S(x, y)$ is the saliency value of the pixel at position (x, y). If the saliency value of a superpixel is larger than threshold, it is considered as the part of salient object. In many applications, high precision and high recall are both required. We thus estimate the F-Measure [5] as:

$$F_{\beta} = \frac{(1 + \beta) \, precision \times recall}{\beta \times precision + recall} \tag{20}$$

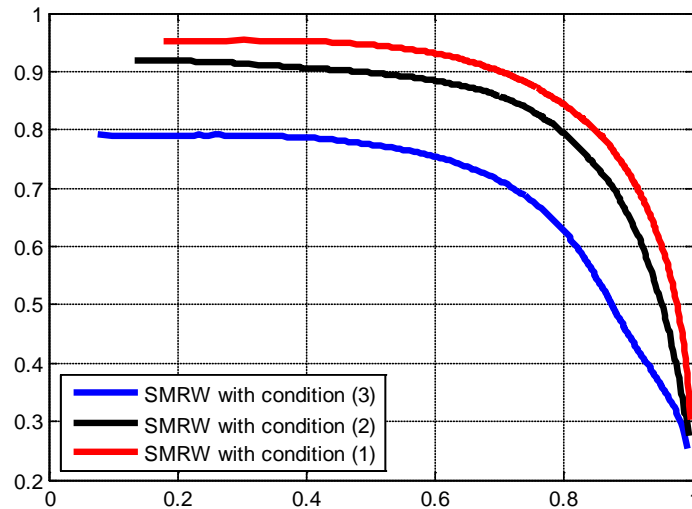where we set $\beta = 0.3$ to emphasize the precision.

Fig. 2. Precision-recall curves of our method with different priors

4.2 Foreground Seeds Selection

As we have discussed the selection of background seeds in section 3.3, the foreground seed is also essential for the proposed algorithm. What is the best seed of salient object in a given image? It obviously depends on the content of the image. However, for the location of salient object is unknown before detection, it is hard to choose the prior of foreground. In order to improve the accuracy of our method, we exploit different ways to decide the seed. For example, (1) we use the center prior, which indicate that the salient object is always raised in the center of the image, to select the superpixel or superpixels located in the center of the image. (2) We consider selecting all of the regions except the background seeds as the salient seeds. (3) We also take all the regions at the four boundaries as background seeds and the rest

regions as foreground seeds to show which one is the best. We test our algorithm with three different seeds selection on MSRA1000 image database respectively. The results are shown in **Fig.2**.
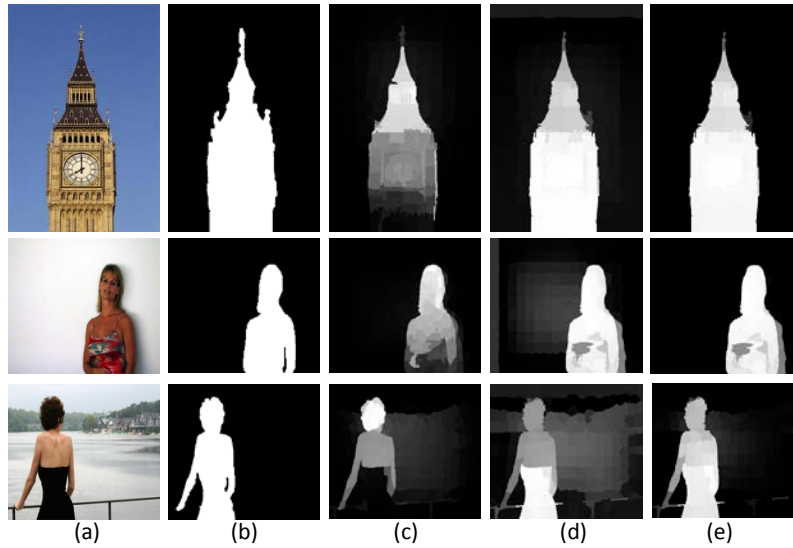

Fig. 3. Some examples. (a) Input image, (b) Ground truth, (c) Saliency maps by prior (3), (d) Saliency maps by prior (2), (e) Saliency maps by prior (1)
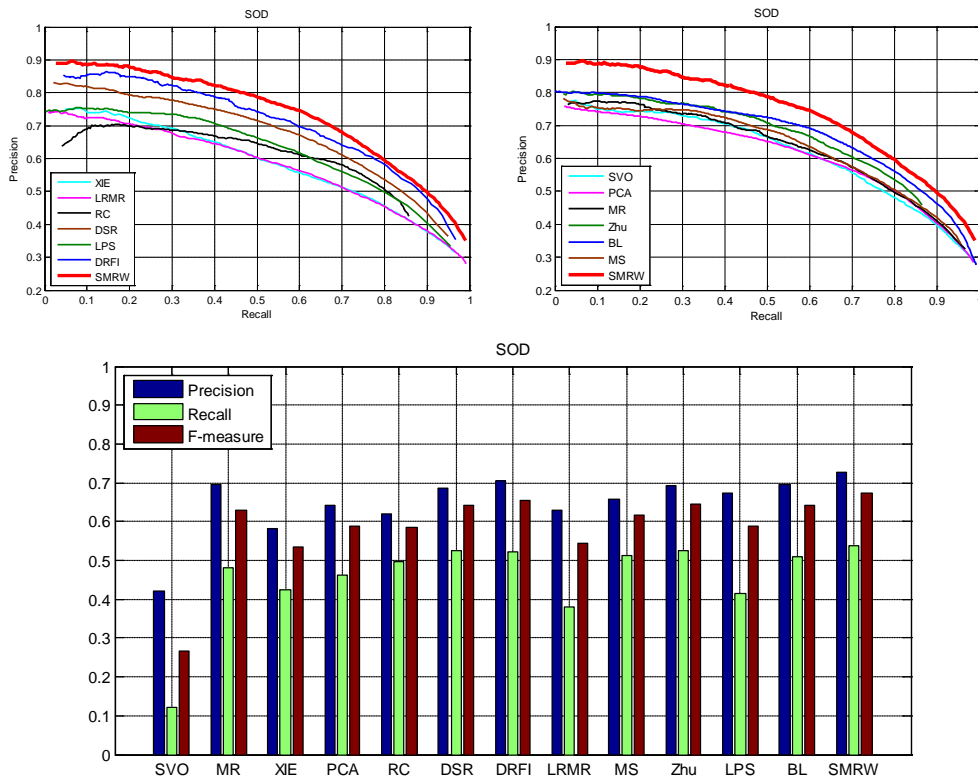

Fig. 4. Quantitative comparisons of saliency maps produced by different approaches on SOD dataset

From **Fig. 2**, we can see that the performance of the proposed algorithm combined with center prior is better than others. Furthermore, it is obvious that eliminating erroneous boundaries before saliency estimation can improve the performance effectively. To analyze the reason specifically, we give some examples, shown in Fig.3, according to different seeds selection. Note that the images in **Fig. 3** have a common Characteristic that the object located in the boundary of the image and the dissimilarity between different parts of an object is very large. When we choose (3) as the seeds selection, it is affected by the seeds, which should be a part of the object and near the boundary of the image, and cannot detect the object as a whole. In **Fig. 3** (c), it can be seen that the proposed algorithm can only detect the part of the object near the center of the image. For (2), it takes the rest regions except boundary prior as the foreground seeds, containing the problematic boundary, and generates better results when compared with (3). The results obtained by (2) are shown in **Fig. 3** (d). The shortcoming of (2) is that it is often affected by the clutter background or the noise of the salient object. Our proposed method takes center prior into considers and generates best results shown in **Fig. 3** (e). Generally speaking, not limit to the strategy of salient seed selection mentioned above, the proposed algorithm can also exploit other method, i.e., key point detection, into our framework to detect salient objects. However, the time consuming of key point detection is larger than center prior, which limit the application of the salient object detection, since the saliency detection is often considered as a preprocessing step in digital image processing. In subsequent experiments, we only use center prior as the foreground seed to complete the salient object detection.
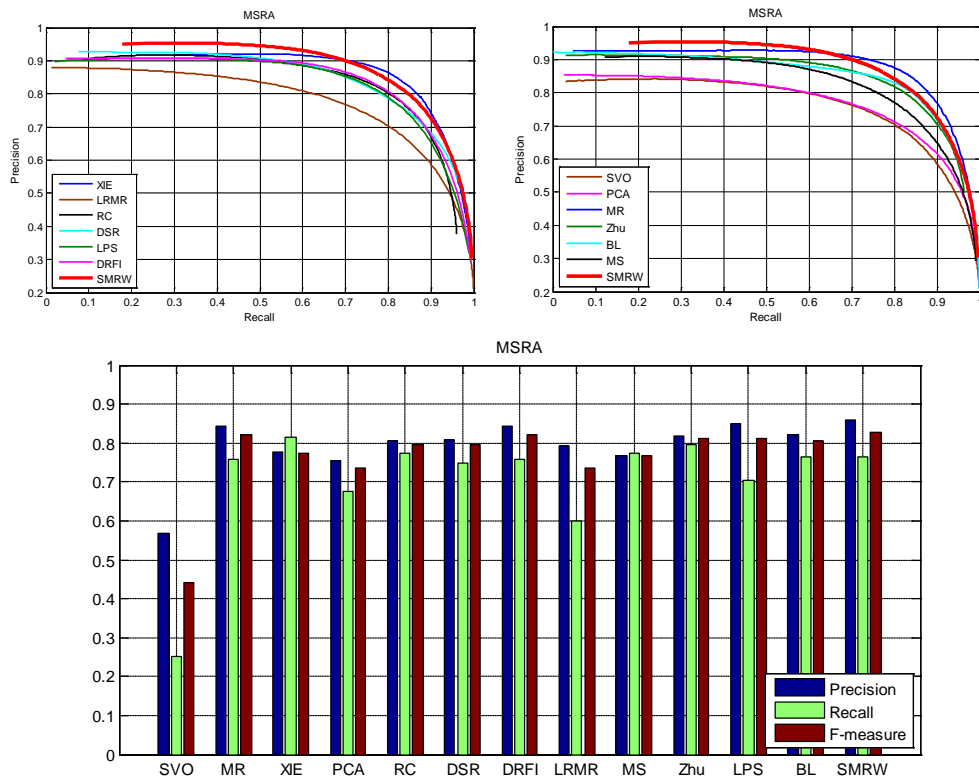

Fig. 5. Quantitative comparisons of saliency maps produced by different approaches on MSRA dataset
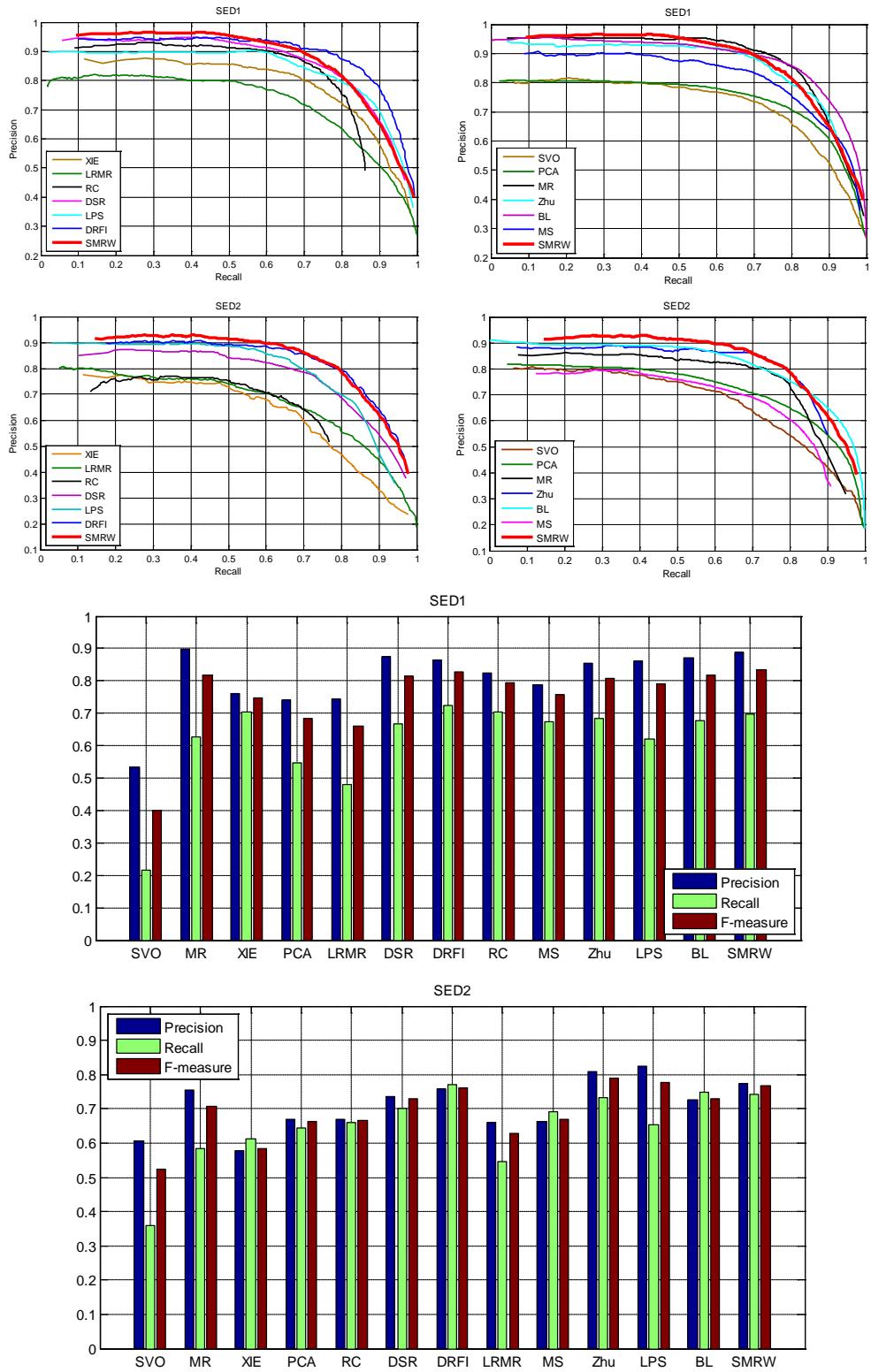
Fig. 6. Quantitative comparisons of saliency maps produced by different approaches on SED dataset

4.3 Comparison to Other Methods

The quantitative comparison is shown in **Fig. 4**-**Fig. 6**. As can be seen, SMRW achieves the best precision on three image datasets. Generally speaking, the precision indicates the performance of the saliency detection algorithms compared with ground-truth saliency map. To compare the proposed model with others, we always see the precision value for different algorithms, for the precision value is the ratio of the correctly detected region over the whole detected region. It slightly improves by 2.18% and 1.02% over the second best algorithms, and 2.83% and 1.64% over the third best algorithms in terms of precision scores according to SOD and MSRA respectively. In addition, SMRW performs well according to precision and recall when compare with MS [44], BL [46], Zhu [11] and LPS [45] in all datasets except SED2. In SED2, the precision of Zhu and LPS is higher than SMRW. Intuitively, our approach has limited ability when discovering all the salient objects within one image (higher recall) or finding the whole part of the salient object in MSRA SED1and SED2 database when compared to DRFI [9]. The reason might be that center prior used in SMRW can highlight the noise which is located in the center of the image. Since the noise is similar to the object, SMRW cannot separate the background and the foreground discriminately on this case. The improvements over state-of-the-arts are substantial when considering their performance and especially the adaptability of our model to different datasets.
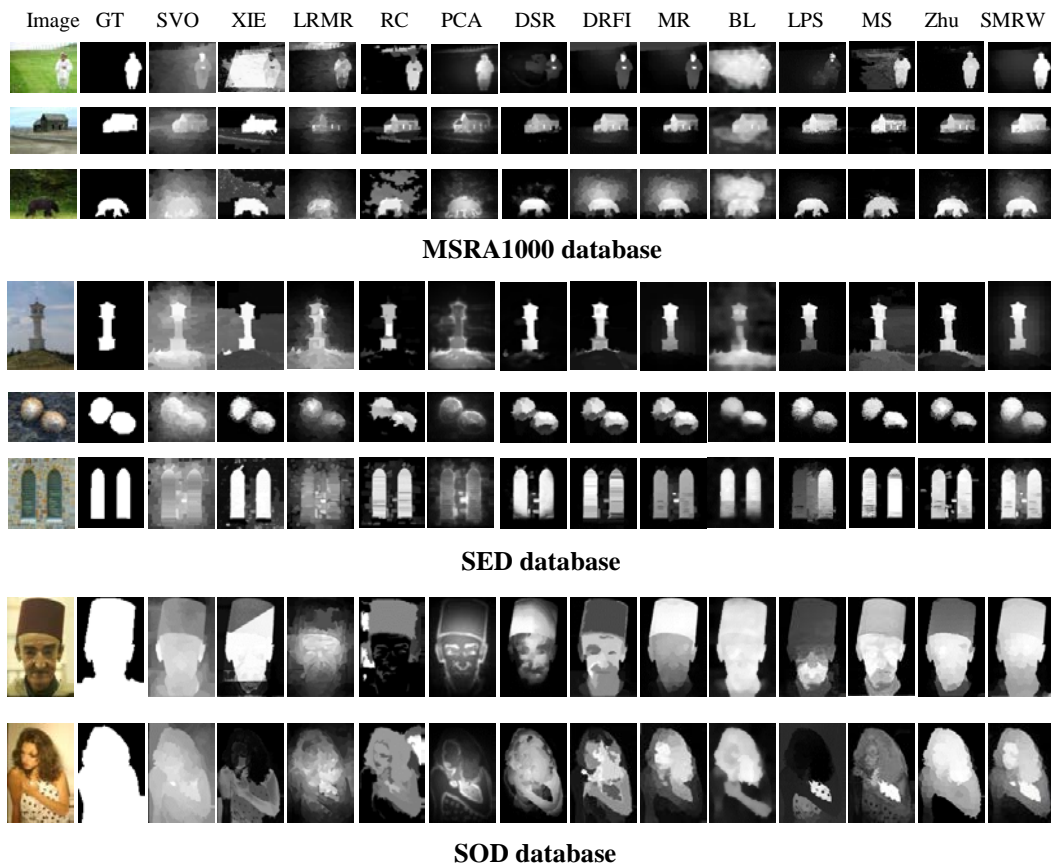

MSRA1000 database

SED database

SOD database

Fig. 7. A part of visual comparisons of the saliency maps

We also provide the visual comparison of different methods in **Fig.7**. As can be seen, our approach can deal well with the challenging cases where the background is cluttered. For example, in the second row, other approaches may be distracted by the textures on the background while our method almost successfully highlights the whole salient object.

It is also worth pointing out that our approach performs well when the object touches the image border, e.g., the third row in MSRA and third row in SED in **Fig.7**, even though it violates the pseudo-background assumption used in MR. MR, which the first stage is based on the pseudo-background assumption, cannot label the saliency seeds correctly when the object touches the image border. LPS [45], which fuses both boundary and objectness labels based on a propagation scheme, exploits the superpixels of one boundary as the seeds of background. Note that it cannot highlight the whole object when the object located at this boundary. Zhu et al. [11] proposed a saliency model which uses robust background measure called boundary connectivity to determine the seeds of background. It can be seen that the saliency maps obtained by Zhu [11] and our algorithm are better than MR and LPS.

For other state-of-the-art approaches, it can be seen that while SVO [40] detects the salient regions, parts of the background are erroneously detected as salient. XIE [43] detects the salient regions that parts of the background are erroneously detected as salient, in which the noise cannot be suppressed in most images. On the contrary, DSR [42] suppress most of the background well while always cannot preserve the object boundary well compared with XIE. It thus generates higher precision scores and less false positive rates. By relying solely on color, RC [6] can mistakenly focus on distinct background colors, e.g., the sky is captured instead of the building in the first row on SOD database. PCA [41] relies mostly on patterns; hence, it detects the outlines of the saliency objects, while missing their interior. LRMR [28], which integrates the high-level priors, focus on the center and the warm color of image. It is worth mentioning that the salient objects with warm colors such as red and yellow are more pronounced. BL [46], which exploits both weak and strong bootstrap learning models, integrate multi-scale saliency maps to improve the detection performance. However, it makes the algorithm cannot suppress the noise in the background and preserve the object boundary well. MS [44] detects the salient object by multi-scale analysis on superpixels. Unlike BL, the results of MS reserve the boundary of salient object better. However, it cannot reserve the object as a whole. For example, in the first line of Fig.7, the face of the girl cannot detect as the salient region. DRFI [9], which is based on multi-level image segmentation, uses the supervised learning approach to map the regional feature vector to a saliency score, and finally fuses the saliency scores across multiple levels, yielding the saliency map. When most of the images contain only one object in training set, it has limited ability to discover all the salient objects within one image.

Table 1. Running time analysis of different methods on different databases (time/s)

Datasets\Methods	SVO	PCA	DRFI	LRMR	MR	MS	LPS	BL	SMRW
MSRA	241.82	39.12	13.97	47.28	1.06	14.68	1.96	58.14	2.46
SOD	229.30	42.48	14.31	58.72	0.967	17.62	2.95	90.82	2.73
SED1	219.95	21.37	11.4	36.96	0.65	9.55	1.27	118.42	2.02
SED2	226.09	15.64	11.48	29.18	0.7	7.34	1.62	141.95	2.18

To compare the running time of the proposed algorithm with other algorithms, we list the time cost in **Table 1**. Here we only list the approaches which use Matlab implementation, i.e., SVO, PCA, LRMR, MR and DRFI, for fair comparison. RC [6], which is implemented by C++ and maintained run-time of 0.6 seconds per image, is faster than the methods aforementioned. However, it is ranked lower in quantitative comparison. We run all algorithms on a machine

with Intel Dual Core E2160 1.8 GHz CPU and 1GB RAM. From **Table 1**, we can see that SVO is much slower than others. The most time-consuming step of SVO is the generation of generic object detector and bottom-up saliency detector. DRFI takes more than 10s for testing a given image, but takes more than 24 hours for training. Our approaches are faster than SVO, PCA, LRMR MS, BL and DRFI. As our approach needs much time to prepare the features, it is slower than MR and LPS but produce superior quality saliency maps than them. The proposed algorithm has relatively high time complexity. The most time consuming part is to calculate the dissimilarity of five different features. It is one of our future research issues to reduce the computational complexity.

5. Conclusion

We propose a bottom-up method to detect salient regions in images based on multiple random walkers. There are two major innovation aspects. Firstly, we propose a novel salient object detection algorithm which is based on multiple random walkers. Under multiple random walks framework, the seeds of background and foreground are considered as the agents to traverse the graph according to a transition matrix. As the process goes on, each walker repels the other walker, while forming a dominant region. Secondly, we design a method of background seeds selection for multiple random walkers in the proposed algorithm. Unlike conventional methods which use a problematic boundary as the seeds of the background in saliency estimation, we optimize the boundary influences by locating and eliminating erroneous boundaries before the saliency detection. We evaluate the proposed algorithm on three different datasets and demonstrate promising results with comparisons to state-of-the-art methods. Our future work will focus on effective salient object seed detection, which could be beneficial for handling more challenging cases and other kinds of high-level knowledge to be embedded into our framework.

References

[1]   A. Toet, "Computational versus Psychophysical Bottom-Up Image Saliency: A Comparative Evaluation Study," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2131-2146, 2011. Article (CrossRef Link).

[2]   A. Borji, D. Sihite, and L. Itti, "Salient object detection: A bench-mark," in *Proc. of ECCV*, pp. 414-429, 2012. Article (CrossRef Link).

[3]   J. Zhou, Z. Jin, "A new framework for multiscale saliency detection based on image patches," *Neural Processing Letters*, vol. 38, no. 3, pp. 361-374, 2013. Article (CrossRef Link).

[4]   L. Itti, C. Koch, E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp.1254–1259, 1998. Article (CrossRef Link).

[5]   R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. of Int. Conf. Computer Vision and Pattern Recognition*, pp.1597-1604, June, 2009. Article (CrossRef Link).

[6]   M. Cheng, N. Mitra, X. Huang, et al. "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp: 409-416, 2015. Article (CrossRef Link).

[7]   K. Fu, C. Gong, J. Yang, et al., "Superpixel based color contrast and color distribution driven salient object detection," *Signal Processing Image Communication*, vol.28, no. 10, pp. 1448-1463, 2013. Article (CrossRef Link).

[8]   L. Mai, Y. Niu, F. Liu, "Saliency Aggregation: A Data-Driven Approach," in *Proc. of Int. Conf. Computer Vision and Pattern Recognition*, pp. 1131-1138, June 23-28, 2013. Article (CrossRef Link).

[9]   H. Jiang, J. Wang, Z. Yuan, et al., "Salient object detection: A discriminative regional feature integration approach," in *Proc. of Int. Conf. Computer Vision and Pattern Recognition*, pp. 2083-2090, June 23-28, 2013. Article (CrossRef Link).

[10] C. Yang, L. Zhang, H. Lu, et al., "Saliency Detection via Graph-Based Manifold Ranking," in *Proc. of Int. Conf. Computer Vision and Pattern Recognition*, pp. 3166-3173, June 23-28, 2013. Article (CrossRef Link).

[11] W. Zhu, S. Liang, Y. Wei, et al., "Saliency Optimization from Robust Background Detection," in *Proc. of Int. Conf. Computer Vision and Pattern Recognition*, pp. 2814-2821, June 23-28, 2014. Article (CrossRef Link).

[12] L. Costa, "Visual saliency and attention as random walks on complex networks," *arXiv preprint*, 2007. Article (CrossRef Link).

[13] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Advances in Neural Information Processing Systems*, pp. 545-552, 2006.

[14] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Random walks on graphs for salient object detection in images," *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3232-3242, 2010. Article (CrossRef Link).

[15] J. Zhou, Y. Ren, Y. Yan, et al., "Salient object detection: manifold-based similarity adaptation approach," *Journal of Electronic Imaging*, vol. 23, no. 6, pp. 063004-063004, 2014. Article (CrossRef Link).

[16] R. Liu, J. Cao, Z. Lin, et al., "Adaptive Partial Differential Equation Learning for Visual Saliency Detection," in *Proc. of Int. Conf. Computer Vision and Pattern Recognition*, pp. 3866-3873, June 23-28, 2014. Article (CrossRef Link).

[17] J. Zhou, S. Gao, Y. Yan, et al., "Saliency detection framework via linear neighborhood propagation," *IET Image Processing*, vol. 8, no. 12, pp. 804-814, 2014. Article (CrossRef Link).

[18] H. Peng, B. Li, W. Xiong, et al., "RGBD Salient Object Detection: A Benchmark and Algorithms," in *Proc. of 13th European Conference Computer Vision (ECCV)*, 2014. Article (CrossRef Link).

[19] H. Peng, B. Li, R. Ji, W. Hu, et al., "Salient Object Detection via Low-Rank and Structured Sparse Matrix Decomposition," in *Proc. of the Twenty-Seventh AAAI Conference on Artificial Intelligence*, pp. 796-802, 2013. Article (CrossRef Link).

[20] J. Yang and M.-H. Yang, "Top-down visual saliency via joint CRF and dictionary learning," in *Proc. of Int. Conf. Computer Vision and Pattern Recognition*, pp. 2296-2303, June 16-21, 2012. Article (CrossRef Link).

[21] A. Borji, D. N. Sihite, and L. Itti, "Probabilistic learning of task-specific visual attention," in *Proc. of Int. Conf. Computer Vision and Pattern Recognition*, pp. 470-477, June 16-21, 2012. Article (CrossRef Link).

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

[23] C. Kanan, M. H. Tong, L. Zhang, and G. W. Cottrell, "SUN: top-down saliency using natural statistics," *Visual Cognition*, vol. 17, no. 6, pp. 979-1003, 2009. Article (CrossRef Link).

[24] L. Elazary, L. Itti, "Interesting objects are visually salient," *Journal of Vision*, vol. 8, no. 3, pp. 1-15, 2008. Article (CrossRef Link).

[25] S. Ramanathan, H. Katti, N. Sebe, et al., "An eye fixation database for saliency detection in images," in *Proc. of 11th European Conference on Computer Vision (ECCV)*, pp. 30-43, 2010. Article (CrossRef Link).

[26] R. Carmi, L. Itti, "Visual causes versus correlates of attentional selection in dynamic scenes," *Vision Research*, vol. 46, no. 26, pp: 4333-4345, 2006. Article (CrossRef Link).

[27] T. Liu, Z. Yuan, J. Sun, et al., "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp: 353–367, 2011. Article (CrossRef Link).

[28] X. Shen, Y. Wu, "A Unified Approach to Salient Object Detection via Low Rank Matrix Recovery," in *Proc. of Int. Conf. Computer Vision and Pattern Recognition*, pp. 853-860, 2012. Article (CrossRef Link).

[29] Y. Jia and M. Han, "Category-independent object-level saliency detection," in *Proc. of Int. Conf. Computer Vision (ICCV)*, pp. 1761-1768, 2013. Article (CrossRef Link).

[30] T. Judd, K. Ehinger, F. Durand, et al., "Learning to predict where humans look," in *Proc. of Int. Conf. Computer Vision (ICCV)*, pp. 2106 – 2113, 2009. Article (CrossRef Link).

[31] H. Jiang, J. Wang, Z. Yuan, et al., "Automatic salient object segmentation based on context and shape prior," in *Proc. of British Machine Vision Conference(BMVC)*, pp: 1-12, 2011. Article (CrossRef Link).

[32] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk, "Salient region detection and segmentation," in *Proc. of 6th Int. Conf. Computer Vision Systems*, pp. 66-75, 2008. Article (CrossRef Link).

[33] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd Edition, Cambridge University Press, 2012.

[34] L. Grady, "Random walks for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1768-1783, 2006. Article (CrossRef Link).

[35] J. Pan, H. Yang, C. Faloutsos, and P. Duygulu, "Automatic multimedia cross-modal correlation discovery," in *Proc. of ACM SIGKDD*, pp. 653-658, 2004. Article (CrossRef Link).

[36] T. Kim, K. Lee, and S. Lee, "Generative image segmentation using random walks with restart," in *Proc. of European Conference on Computer Vision*, pp. 264-275, October 12-18, 2008. Article (CrossRef Link).

[37] J. Kim, J. Sim, C. Kim, "Multiscale Saliency Detection Using Random Walk With Restart," *IEEE Transactions on Circuits & Systems for Video Technology*, vol. 24, no. 2, pp.198-210, 2014. Article (CrossRef Link).

[38] H. Kim, Y. Kim, J. Sim, et al., "Spatiotemporal Saliency Detection for Video Sequences Based on Random Walk with Restart," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2552-2564, 2015. Article (CrossRef Link).

[39] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proc. of Int. Conf. Computer Vision (ICCV)*, pp. 1470-1477, October 13-16, 2003. Article (CrossRef Link).

[40] K. Chang, T. Liu, H. Chen, and S. Lai, "Fusing generic objectness and visual saliency for salient object detection," in *Proc. of Int. Conf. Computer Vision (ICCV)*, pp. 914-921, Nov. 6-13, 2011. Article (CrossRef Link).

[41] R. Margolin, A. Tal and L. Zelnik-Manor, "What Makes a Patch Distinct?" in *Proc. of Int. Conf. Computer Vision and Pattern Recognition*, pp.1139-1146, June 23-28, 2013. Article (CrossRef Link).

[42] X. Li, H. Lu, L. Zhang, X. Ruan, M. Yang, "Saliency Detection via Dense and Sparse Reconstruction," in *Proc. of Int. Conf. Computer Vision (ICCV)*, pp. 2976-2983, Dec. 1-8, 2013. Article (CrossRef Link).

[43] Y. Xie, H. Lu, and M. Yang, "Bayesian saliency via low and mid-level cues," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1689-1698, 2013. Article (CrossRef Link).

[44] N. Tong, H. Lu, L. Zhang, X. Ruan, "Saliency Detection with Multi-Scale Superpixels," *IEEE Signal Processing Letters*, vol. 21, no. 9, pp. 1035-1039, 2014. Article (CrossRef Link).

[45] H. Li, H. Lu, Z. Lin, et al., "Inner and Inter Label Propagation: Salient Object Detection in the Wild," *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 3176-3186, 2015. Article (CrossRef Link).

[46] N. Tong, H. Lu, L. Zhang, X. Ruan., M. Yang, "Salient object detection via bootstrap learning," in *Proc. of Int. Conf. Computer Vision and Pattern Recognition*, pp. 1884-1892, 2015. Article (CrossRef Link).

[47] B. Alexe, T. Deselaers, and V. Ferrari, "What is an object?" in *Proc. of Int. Conf. Computer Vision and Pattern Re*cognition, pp. 73-80, June13-18, 2010. Article (CrossRef Link).

[48] S. Alpert, M. Galun, R. Basri, and A. Brandt, "Image segmentation by probabilistic bottom-up aggregation and cue integration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 315-327, 2011. Article (CrossRef Link).

Jiyou Zhai is a doctoral student majoring in College of Computer and Information at Hohai University from September 2012 to present. He is a lecturer at Nanjing Institute of Technology. His research interests include pattern recognition, image processing, etc.

Jingbo Zhou received his PhD degree in control science and engineering from Nanjing University of Science and Technology (NUST) in 2013. He is a lecturer at Huaiyin Institute of Technology. His research interests include pattern recognition, image processing, data clustering, etc.

Yongfeng Ren is a doctoral student majoring in College of Computer and Information at Hohai University from September 2012 to present. He is a lecturer at Huaiyin Institute of Technology. His research interests include pattern recognition, image processing, etc.

Zhijian Wang received his PhD degree in computer software from Nanjing University. He is a professor in College of Computer and Information, Hohai University. His research interests include pattern recognition, computer software, etc.