Saliency Detection via Absorbing Markov Chain

Bowen Jiang[1], Lihe Zhang[1], Huchuan Lu[1], Chuan Yang[1], and Ming-Hsuan Yang[2]
[1]Dalian University of Technology [2]University of California at Merced

Abstract

In this paper, we formulate saliency detection via absorbing Markov chain on an image graph model. We jointly consider the appearance divergence and spatial distribution of salient objects and the background. The virtual boundary nodes are chosen as the absorbing nodes in a Markov chain and the absorbed time from each transient node to boundary absorbing nodes is computed. The absorbed time of transient node measures its global similarity with all absorbing nodes, and thus salient objects can be consistently separated from the background when the absorbed time is used as a metric. Since the time from transient node to absorbing nodes relies on the weights on the path and their spatial distance, the background region on the center of image may be salient. We further exploit the equilibrium distribution in an ergodic Markov chain to reduce the absorbed time in the long-range smooth background regions. Extensive experiments on four benchmark datasets demonstrate robustness and efficiency of the proposed method against the state-of-the-art methods.

1. Introduction

Saliency detection in computer vision aims to find the most informative and interesting region in a scene. It has been effectively applied to numerous computer vision tasks such as content based image retrieval [32] , image segmentation [30], object recognition [24] and image adaptation [21]. Existing methods are developed with bottom-up visual cues [19, 10, 26, 34] or top-down models [4, 36].

All bottom-up saliency methods rely on some prior knowledge about salient objects and backgrounds, such as contrast, compactness, etc. Different saliency methods characterize the prior knowledge from different perspectives. Itti *et al.* [16] extract center-surround contrast at multiple spatial scales to find the prominent region. Bruce *et al.* [6] exploit Shannons self-information measure in local context to compute saliency. However, the local contrast does not consider the global influence and only stands out at object boundaries. Region contrast based methods [8, 17] first segment the image and then compute the global contrast of those segments as saliency, which can usually highlight the entire object. Fourier spectrum analysis has also been used to detect visual saliency [15, 13]. Recently, Perazzi *et al.* [25] unify the contrast and saliency computation into a single high-dimensional Gaussian filtering framework. Wei *et al.* [33] exploit background priors and geodesic distance for saliency detection. Yang *et al.* [35] cast saliency detection into a graph-based ranking problem, which performs label propagation on a sparsely connected graph to characterize the overall differences between salient object and background.

In this work, we reconsider the properties of Markov random walks and their relationship with saliency detection. Existing random walk based methods consistently use the equilibrium distribution in an ergodic Markov chain [9, 14] or its extensions, e.g. the site entropy rate [31] and the hitting time [11], to compute saliency, and have achieved success in their own aspects. However, these models still have some certain limitations. Typically, saliency measure using the hitting time often highlights some particular small regions in objects or backgrounds. In addition, equilibrium distribution based saliency models only highlight the boundaries of salient object while object interior still has low saliency value. To address these issues, we investigate the properties of absorbing Markov chains in this work. Given an image graph as Markov chain and some absorbing nodes, we compute the expected time to absorption (i.e. the absorbed time) for each transient node. The nodes which have similar appearance (i.e. large transition probabilities) and small spatial distance to absorbing nodes can be absorbed faster. As salient objects seldom occupy all four image boundaries [33, 5] and the background regions often have appearance connectivity with image boundaries, when we use the boundary nodes as absorbing nodes, the random walk starting in background nodes can easily reach the absorbing nodes. While object regions often have great contrast to the image background, it is difficult for a random walk from object nodes to reach these absorbing nodes (represented by boundary nodes). Hence, the absorbed time starting from object nodes is longer than that from background nodes. In addition, in a long run, the absorbed time with similar starting nodes is roughly the same. Inspired

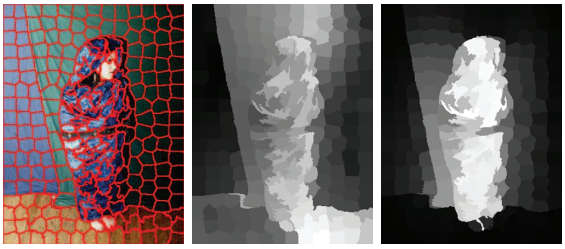

Figure 1. The time property of absorbing Markov chain and ergodic Markov chain. From left to right: input image with superpixels as nodes; the minimum hitting time of each node to all boundary nodes in ergodic Markov chain; the absorbed time of each node into all boundary nodes in absorbing Markov chain. Each kind of time is normalized as a saliency map respectively.

by these observations, we formulate saliency detection as a random walk problem in the absorbing Markov chain.

The absorbed time is not always effective especially when there are long-range smooth background regions near the image center. We further explore the effect of the equilibrium probability in saliency detection, and exploit it to regulate the absorbed time, thereby suppressing the saliency of this kind of regions.

2. Related Work

Previous works that simulate saliency detection in random walk model include [9, 14, 11, 31]. Costa *et al.* [9] identify the saliency region based on the frequency of visits to each node at the equilibrium of the random walk. Harel *et al.* [14] extend the above method by defining a dissimilar measure to model the transition probability between two nodes. In [31], Wang *et al.* introduce the entropy rate and incorporate the equilibrium distribution to measure the average information transmitted from a node to the others at one step, which is used to predict visual attention. A major problem using the equilibrium distribution is that this approach often only highlights the texture and boundary regions rather than the entire object, as the equilibrium probability in the cluttered region is larger than in homogeneous region when using the dissimilarity of two nodes to represent their transition probability. Furthermore, the main objectives in [9, 14, 31] are to predict human fixations on natural images as opposed to identifying salient regions that correspond to objects, as illustrated in this paper.

The approach most related to ours is Gopalakrishnan *et al.* [11], which exploits the hitting time on the fully connected graph and the sparsely connected graph to find the most salient seed, based on which some background seeds are determined again. They then use the difference of the hitting times to the two kinds of seeds to compute the saliency for each node. While they alleviate the problem of using the equilibrium distribution to measure saliency, the identification of the salient seed is difficult, especially for the scenes with complex salient objects. More importantly, the hitting time based saliency measure prefers to highlight the global rare regions and does not suppress the backgrounds very well, thereby decreasing the overall saliency of objects (See Figure 1). This can be explained as follows. The hitting time is the expected time taken to reach a node if the Markov chain is started in another node. The ergodic Markov chain doesn't have a mechanism that can synthetically consider the relationships between a node and multiple specific nodes (e.g. seed nodes). In [11], to describe the relevance of a node to background seeds, they use the minimum hitting time to take all the background seeds into account. The minimum time itself is sensitive to some noise regions in the image.

Different from the above methods, we consider the absorbing Markov random walk, which includes two kinds of nodes (i.e. absorbing nodes and transient nodes), to measure saliency. For an absorbing chain started in a transient node, the probability of absorption in an absorbing node indicates the relationship between the two nodes, and the absorption time therefore implicates the selective relationships between this transient node and all the absorbing nodes. Since the boundary nodes usually contain the global characteristics of the image background, by using them as absorbing nodes, the absorbed time of each transient node can reflect its overall similarity with the background, which helps to distinguish salient nodes from background nodes. Moreover, as the absorbed time is the expected time to all the absorbing nodes, it covers the effect of all the boundary nodes, which can alleviate the influence of particular regions and encourage the similar nodes in a local context to have the similar saliency, thereby overcoming the defects of using the equilibrium distribution [9, 14, 11, 31]. Different from [9, 14] which directly use the equilibrium distribution to simulate human attention, we exploit it to weigh the absorbed time, thereby suppressing the saliency of long-range background regions with homogeneous appearance.

3. Principle of Markov Chain

Given a set of states $S = \{s_1, s_2, \ldots, s_m\}$, a Markov chain can be completely specified by the $m \times m$ transition matrix $\mathbf{P}$, in which p_{ij} is the probability of moving from state s_i to state s_j. This probability does not depend upon which state the chain is in before the current state. The chain starts in some state and move from one state to another successively.

3.1. Absorbing Markov Chain

The state s_i is absorbing when $p_{ii} = 1$, which means $p_{ij} = 0$ for all $i \neq j$. A Markov chain is absorbing if it has at least one absorbing state. It is possible to go from every transient state to some absorbing state, not necessarily in one step. Considering an absorbing chain with r absorbing

states and t transient states, renumber the states so that the transient states come first, then the transition matrix $\mathbf{P}$ has the following canonical form,

$$\mathbf{P} \rightarrow \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}, \tag{1}$$

where the first t states are transient and the last r states are absorbing. $\mathbf{Q} \in [0,1]^{t \times t}$ contains the transition probabilities between any pair of transient states, while $\mathbf{R} \in [0,1]^{t \times r}$ contains the probabilities of moving from any transient state to any absorbing state. $\mathbf{0}$ is the $r \times t$ zero matrix and $\mathbf{I}$ is the $r \times r$ identity matrix.

For an absorbing chain, we can derive its fundamental matrix $\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1}$, where n_{ij} can be interestingly interpreted as the expected number of times that the chain spends in the transient state j given that the chain starts in the transient state i , and the sum $\sum_j n_{ij}$ reveals the expected number of times before absorption (into any absorbing state). Thus, we can compute the absorbed time for each transient state, that is,

$$\mathbf{y} = \mathbf{N} \times \mathbf{c}, \tag{2}$$

where $\mathbf{c}$ is a t dimensional column vector all of whose elements are 1.

3.2. Ergodic Markov Chain

An ergodic Markov chain is one in which it is possible to go from every state to every state, not necessarily in one step. An ergodic chain with any starting state always reaches equilibrium after a certain time, and the equilibrium state is characterized by the equilibrium distribution π, which satisfies the equation

$$\pi \mathbf{P} = \pi, \tag{3}$$

where $\mathbf{P}$ is the ergodic transition matrix. π is a strictly positive probability vector, where π_i describes the expected probability of the chain staying in state s_i in equilibrium. When the chain starts in state s_i, the mean recurrent time h_i (i.e., the expected number of times to return to state s_i) can be derived from the equilibrium distribution π. That is,

$$h_i = \frac{1}{\pi_i}, \tag{4}$$

where i indexes all the states in the ergodic Markov chain. The more states there are similar to state s_i nearby, the less h_i is. The derivation details and proofs can be found in [12].

3.3. Saliency Measure

Given an input image represented as a Markov chain and some background absorbing states, the saliency of each transient state is defined as the expected number of times before being absorbed into all absorbing nodes by Eq 2. In this work, the transition matrix is constructed on a sparsely connected graph, where each node corresponds to a state. Because we compute the full resolution saliency map, some virtual nodes are added to the graph as absorbing states, which is detailed in the next section.

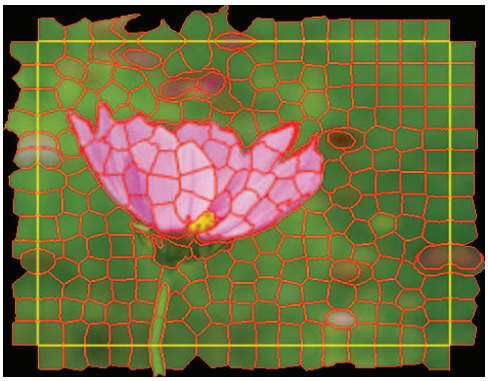

Figure 2. Illustration of the absorbing nodes. The superpixels outside the yellow bounding box are the duplicated boundary superpixels, which are used as the absorbing nodes.

In the conventional absorbing Markov chain problems, the absorbing nodes are manually labelled with the ground-truth. However, as absorbing nodes for saliency detection are selected by the proposed algorithm, some of them may be incorrect. They have insignificant effects on the final results, which are explained in the following sections.

4. Graph Construction

We construct a single layer graph $G(V, E)$ with superpixels [3] as nodes V and the links between pairs of nodes as edges E. Because the salient objects seldom occupy all image borders [33], we duplicate the boundary superpixels around the image borders as the virtual background absorbing nodes, as shown in Figure 2. On this graph, each node (transient or absorbing) is connected to the transient nodes which neighbour it or share common boundaries with its neighbouring nodes. That means that any pair of absorbing nodes are unconnected. In addition, we enforce that all the transient nodes around the image borders (i.e., boundary nodes) are fully connected with each other, which can reduce the geodesic distance between similar superpixels. The weights of the edges encode nodal affinity such that nodes connected by an edge with high weight are considered to be strongly connected and edges with low weights represent nearly disconnected nodes. In this work, the weight w_{ij} of the edge e_{ij} between adjacent nodes i and j is defined as

$$w_{ij} = e^{-\frac{\|x_i - x_j\|}{\sigma^2}}, \tag{5}$$

where x_i and x_j are the mean of two nodes in the CIE LAB color space, and σ is a constant that controls the strength of

the weight. We first renumber the nodes so that the first t nodes are transient nodes and the last r nodes are absorbing nodes, then define the affinity matrix $\mathbf{A}$ which represents the reverence of nodes as

$$a_{ij} = \begin{cases} w_{ij} & j \in N(i), 1 \leq i \leq t \\ 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where $N(i)$ denotes the nodes connected to node i. The degree matrix that records the sum of the weights connected to each node is written as

$$\mathbf{D} = \text{diag}(\sum\nolimits_j a_{ij}). \quad (7)$$

Finally, the transition matrix $\mathbf{P}$ on the sparsely connected graph is given as

$$\mathbf{P} = \mathbf{D}^{-1} \times \mathbf{A}, \quad (8)$$

which is actually the raw normalized $\mathbf{A}$. As the nodes are locally connected, $\mathbf{P}$ is a sparse matrix with a small number of nonzero elements.

The sparsely connected graph restricts the random walk to only move within a local region in each step, hence the expected time spent to move from transient node v_t to absorbing node v_a is determined by two major factors. One is the spatial distance between the two nodes. Their distance is larger, and the expected time is longer. The other is the transition probabilities of the nodes along the different paths from v_t to v_a. Large probabilities are able to shorten the expected time to absorption. Given starting node v_t, the shorter the time is, the larger the probability of absorption in node v_a is in a long run.

5. Saliency Detection

Given the transition matrix $\mathbf{P}$ by Eq. 8, we can easily extract the matrix $\mathbf{Q}$ by Eq. 1, based on which the fundamental matrix $\mathbf{N}$ is computed. Then, we obtain the saliency map $\mathbf{S}$ by normalizing the absorbed time $\mathbf{y}$ computed by Eq. 2 to the range between 0 and 1, that is

$$\mathbf{S}(i) = \overline{\mathbf{y}}(i) \qquad i = 1, 2, \ldots, t, \quad (9)$$

where i indexes the transient nodes on graph, and $\overline{\mathbf{y}}$ denotes the normalized absorbed time vector.

Most saliency maps generated by the normalized absorbed time $\overline{\mathbf{y}}$ are effective, but some background nodes near the image center may not be adequately suppressed when they are in long-range homogeneous region, as shown in Figure 3. That can be explained as follows. Most nodes in this kind of background regions have large transition probabilities, which means that the random walk may transfer many times among these nodes before reaching the

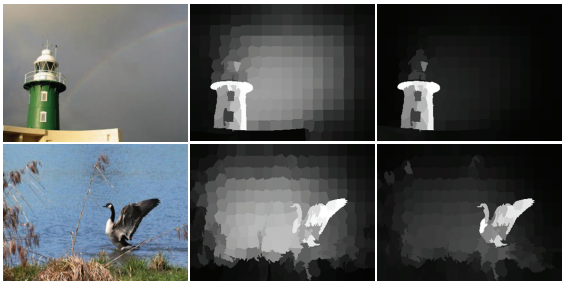

Figure 3. Examples showing the benefits of the update processing. From left to right, input images, results without and with the update processing.

absorbing nodes. The sparse connectivity of the graph results that the background nodes near the image center have longer absorbed time than the similar nodes near the image boundaries. Consequently, the background regions near the image center possibly present comparative saliency with salient objects, thereby decreasing the contrast of objects and backgrounds in the resulted saliency maps. To alleviate this problem, we update the saliency map by using a weighted absorbed time $\mathbf{y_w}$, which can be denoted as:

$$\mathbf{y_w} = \mathbf{N} \times \mathbf{u}, \quad (10)$$

where $\mathbf{u}$ is the weighting column vector. In this work, we use the normalized recurrent time of an ergodic Markov chain, of which the transition matrix is the row normalized $\mathbf{Q}$, as the weight $\mathbf{u}$.

The equilibrium distribution π for the ergodic Markov chain can be computed from the affinity matrix $\mathbf{A}$ as

$$\pi_i = \frac{\sum_j a_{ij}}{\sum_{ij} a_{ij}}, \quad (11)$$

where i, j index all the transient nodes. Since we define the edge weight w_{ij} as the similarity between two nodes, the nodes within the homogeneous region have large weighted sum $\sum_j a_{ij}$. This means the recurrent time in this kind of region is small as shown in Figure 3. For this reason, we use the average recurrent time h_j of each node j to weight the corresponding element n_{ij} (i.e., the expected time spending in node j before absorption given starting node i) in each row of the fundamental matrix $\mathbf{N}$. Precisely, given the equilibrium distribution π, h_j is computed by Eq. 4 and the weighting vector $\mathbf{u}$ is computed as:

$$u_j = \frac{h_j}{\sum_k h_k}, \quad (12)$$

where j and k index all the transient nodes on graph.

By the update processing, the saliency of the long-range homogeneous regions near the image center can be suppressed as Figure 3 illustrates. However, if the kind of region belongs to salient object, its saliency will be also incorrectly suppressed. Therefore, we define a principle to

Figure 4. Examples in which the salient objects appear at the image boundaries. From top to down: input images, our saliency maps.

decide which maps need to be further updated. We find that object regions have great global contrast to background regions in good saliency maps, while it is not the case in the defective maps as the examples in Figure 3, which consistently contain a number of regions with mid-level saliency. Hence, given a saliency map, we first calculate its gray histogram $\mathbf{g}$ with ten bins, and then define a metric $score$ to characterize this kind of tendency as follows:

$$score = \sum_{b=1}^{10} g(b) \times \min(b, (11 - b)), \tag{13}$$

where b indexes all the bins. The larger $score$ means that there are longer-range regions with mid-level saliency in the saliency map.

It should be noted that the absorbing nodes may include object nodes when the salient objects touch the image boundaries, as shown in Figure 4. These imprecise background absorbing nodes may result that the object regions close to the boundary are suppressed. However, the absorbed time considers the effect of all boundary nodes and depends on two factors: the edge weights on the path and the spatial distance, so the parts of object which are far from or different from the boundary absorbing nodes can be highlighted correctly. The main procedure of the proposed method is summarized in Algorithm 1.

Algorithm 1 Saliency detection based on Markov random walk

Input: An image and required parameters.

1. Construct a graph G with superpixels as nodes, and use boundary nodes as absorbing nodes;
2. Compute the affinity matrix $\mathbf{A}$ by Eq. 6 and the transition matrix $\mathbf{P}$ by Eq. 8;
3. Extract the matrix $\mathbf{Q}$ from $\mathbf{P}$ by Eq. 1, and compute the fundamental matrix $\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1}$ and the map $\mathbf{S}$ by Eq. 9;
4. Compute the $score$ by Eq. 13, if $score < \gamma$, output $\mathbf{S}$ and return;
5. Compute the recurrent time $\mathbf{h}$ by Eq. 11 and 4, and the weight $\mathbf{u}$ by Eq. 12, then compute the saliency map $\mathbf{S}$ by Eq. 10 and 9;

Output: the full resolution saliency map.

6. Experimental Results

We evaluate the proposed method on four benchmark datasets. The first one is the MSRA dataset [18] which contains 5,000 images with the ground truth marked by bounding boxes. The second one is the ASD dataset, a subset of the MSRA dataset, which contains 1,000 images with accurate human-labelled ground truth provided by [2]. The third one is the SED dataset [28], which contains: the single object sub-dataset SED1 and two objects sub-dataset SED2. Each sub-dataset contains 100 images and have accurate human-labelled ground truth. The fourth one is the most challenging SOD dataset which contains 300 images from the Berkeley segmentation dataset [22]. This dataset is first used for salient object segmentation evaluation [23], where seven subjects are asked to label the foreground salient object masks. For each object mask of each subject, a consistency score is computed based on the labels of the other six subjects. We select and combine the object masks whose consistency scores are higher than 0.7 as the final ground truth as done in [33]. We compare our method with fifteen state-of-the-art saliency detection algorithms: the IT [16], MZ [20], LC [37], GB [14], SR [15], AC [1], FT [2], S-ER [31], CA [27], RC [8], CB [17], SVO [7], SF [25], L-R [29] and GS [33] methods.

Experimental Setup: We set the number of superpixel nodes $N = 250$ in all the experiments. There are two parameters in the proposed algorithm: the edge weight σ in Eq. 5 to controls the strength of weight between a pair of nodes and the threshold γ of $score$ in Eq. 13 to indicate the quality of the saliency map. These two parameters are empirically chosen, $\sigma^2 = 0.1$ and $\gamma = 2$ for all the test images in the experiments.

Evaluation Metrics: We evaluate all methods by precision, recall and F-measure. The precision is defined as the ratio of salient pixels correctly assigned to all the pixels of extracted regions. The recall is defined as the ratio of detected salient pixels to the ground-truth number. The F-measure is the overall performance measurement computed by the weighted harmonic of precision and recall:

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

We set $\beta^2 = 0.3$ to stress precision more than recall, the same to [2, 8, 25]. Similar as previous works, two evaluation criteria are used in our experiments. First, we bi-segment the saliency map using every threshold in the range $[0 : 0.05 : 1]$, and compute precision and recall at each value of the threshold to plot the precision-recall curve. Second, we compute the precision, recall and F-measure with an adaptive threshold proposed in [2], which is defined as twice the mean saliency of the image.

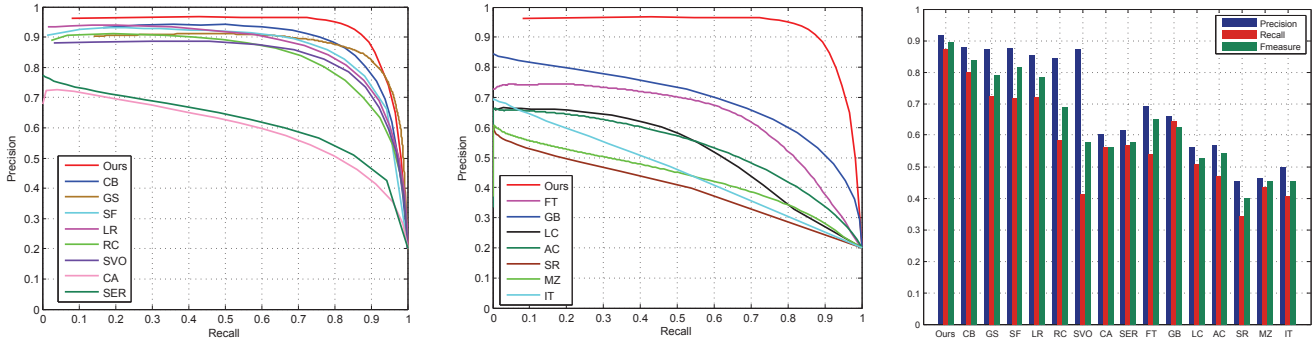

Figure 5. Evaluation results on the ASD dataset. Left, middle: precision and recall rates for all algorithms. Right: precision, recall, and F-measure for adaptive thresholds. Our approach consistently outperforms all other methods.

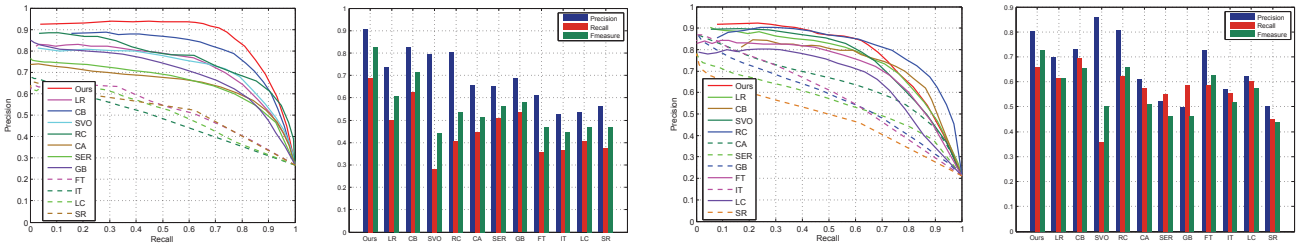

Figure 6. Evaluation results on the SED dataset. Left two: the results for different methods on the SED1 dataset. Right two: the results for different methods on the SED2 dataset.

ASD: We evaluate the performance of the proposed method against fifteen state-of-the-art methods. The results are shown in Figure 5. The two evaluation criteria consistently show the proposed method outperforms all the other methods, where the CB [17], SVO [7], RC [8] and CA [27] are top-performance methods for saliency detection in a recent benchmark study [5]. Some visual comparison examples are shown in Figure 9 and more results can be found in the supplementary material. We note that the proposed method more uniformly highlights the salient regions while adequately suppresses the backgrounds than the other methods.

MSRA: On the MSRA dataset, we compare the proposed method with eleven state-of-the-art methods which are LR [29], CB [17], SVO [7], RC [8], CA [27], SER [31], FT [2], GB [14], SR [15], LC [37] and IT [16]. This dataset contains the ground truth of salient region marked as bounding boxes by nine subjects. We accumulate the nine ground truth, and then choose the pixels with consistency score higher than 0.5 as salient region and fit a bounding box in the salient region. The bounding box is the final ground truth. Similar as previous works, we first fit a rectangle to the binary saliency map and then use the bounding box to compute precision, recall and F-measure. Figure 7 shows that the proposed method performs better than the other methods on this large dataset. The SVO method has larger precision value, since it tends to detect the most salient regions at the expense of low recall. While the CA [27], IT [16], FT [2], SR [15] and LC [37] methods also show the same imbalance. Their recalls for adaptive thresholds are quite high and close to 1. That is because the background is suppressed badly, the cut saliency map contains almost the entire image with low precision.

SED: On this single object and two object dataset, we compare the proposed method with eleven state-of-the-art methods which are LR [29], CB [17], SVO [7], RC [8], CA [27], SER [31], FT [2], GB [14], SR [15], LC [37] and IT [16]. As shown in Figure 6, the proposed method performs best on the SED1 dataset, while performs poorly compared with the RC and CB methods at the recall values from 0.7 to 1 on the SED2 dataset. That is because our method usually highlights one of two objects while the other has low saliency values due to the appearance diversity of two objects.

SOD: On this most challenging dataset, we evaluate the performance of the post-process step against the map obtained directly form absorbed time (noted 'Before') and twelve state-of-the-art methods as shown in Figure 7. We can see that the post-process step improves the precision and recall significantly over the solely saliency measure by absorbed time. The two evaluation criteria show the proposed method performs equally well or slightly better than the GS [33] method. Some visual examples are given in Figure 9. Due to scrambled backgrounds and heterogeneous foregrounds most images have, and the lack of top-down prior knowledge, the overall performance of the existing bottom-up saliency detection methods is low on this dataset.

Failure Case: Our approach exploits the boundary prior to determine the absorbing nodes, therefore the small salient object touching image boundaries may be incorrectly suppressed. According to the computation of the absorbed

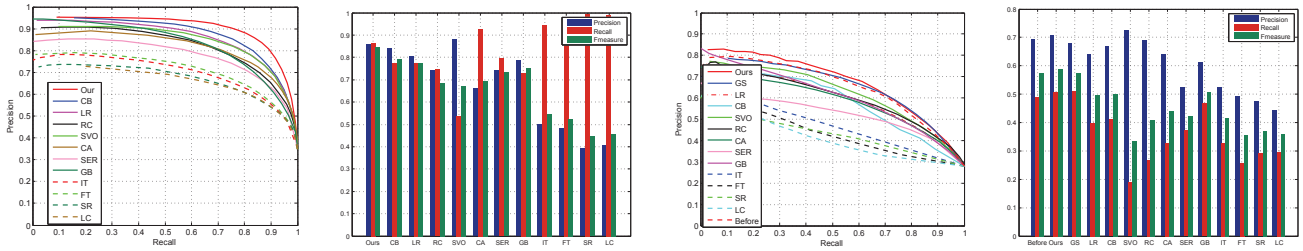

Figure 7. Evaluation results on the MSRA and SOD dataset. Left two: the results for different methods on the MSRA dataset. Right two: the results for different methods on the SOD dataset.

Figure 8. Failure examples

time, a node with sharp contrast to its surroundings often has abnormally large absorbed time, which results that most parts of object even the whole object are suppressed. In addition, the object with similar appearance to the background is very difficult to be detected, which is a known problem in object detection. Figure 8 shows the typical failure cases.

Execution Time: Generally, better results can be achieved at the expense of execution time. We compare the execution time of different methods. The average execution time of state-of-the-art methods are summarized in Table 1 on an Intel i7 3.40GHz CPU with 32GB RAM. Our Matlab implementation is available at `http://ice.dlut.edu.cn/lu/publications.html`, or `http://faculty.ucmerced.edu/mhyang/pubs.html`.

7. Conclusion

In this paper, we propose a bottom-up saliency detection algorithm by using the time property in an absorbing Markov chain. Based on the boundary prior, we set the virtual boundary nodes as absorbing nodes. The saliency of each node is computed as its absorbed time to absorbing nodes. Furthermore, we exploit the equilibrium distribution in ergodic Markov chain to weigh the absorbed time, thereby suppressing the saliency in long-range background regions. Experimental results show that the proposed method outperforms fifteen state-of-the-art methods on the four public datasets and is computationally efficient.

Acknowledgement

B. Jiang, L. Zhang and C. Yang are supported by the Natural Science Foundation of China #61371157 and the Fundamental Research Funds for the Central Universities (DUT12JS05). H. Lu is supported by the Natural Science Foundation of China #61071209 and #61272372. M.-H. Yang is supported by the NSF CAREER Grant #1149783 and NSF IIS Grant #1152576.

References

[1] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk. Salient region detection and segmentation. In *ICVS*, 2008.

[2] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In *CVPR*, 2009.

[3] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. Slic superpixels. *Technical Report, EPFL*, 2010.

[4] A. Borji. Boosting bottom-up and top-down visual features for saliency estimation. In *CVPR*, 2012.

[5] A. Borji, D. N. Sihite, and L. Itti. Salient object detection: A benchmark. In *ECCV*, 2012.

[6] N. Bruce and J. Tsotsos. Saliency based on information maximization. *NIPS*, 2006.

[7] K. Chang, T. Liu, H. Chen, and S. Lai. Fusing generic objectness and visual saliency for salient object detection. In *ICCV*, 2011.

[8] M. Cheng, G. Zhang, N. J. Mitra, X. Huang, and S. Hu. Global contrast based salient region detection. In *CVPR*, 2011.

[9] L. d. F. Costa. Visual saliency and attention as random walks on complex networks. *arXiv preprint physics/0603025*, 2006.

[10] D. Gao and N. Vasconcelos. Bottom-up saliency is a discriminant process. In *ICCV*, 2007.

[11] V. Gopalakrishnan, Y. Hu, and D. Rajan. Random walks on graphs for salient object detection in images. *TIP*, 2010.

[12] C. M. Grinstead and J. L. Snell. *Introduction to probability*. American Mathematical Soc., 1998.

[13] C. Guo, Q. Ma, and L. Zhang. Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In *CVPR*, 2008.

[14] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *NIPS*, 2006.

[15] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In *CVPR*, 2007.

[16] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *PAMI*, 1998.

[17] 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.

[18] T. Liu, J. Sun, N. Zheng, X. Tang, and H.-Y. Shum. Learning to detect a salient object. In *CVPR*, 2007.

[19] Y. Lu, W. Zhang, H. Lu, and X. Xue. Salient object detection using concavity context. In *ICCV*, 2011.

[20] Y. Ma and H. Zhang. Contrast-based image attention analysis by using fuzzy growing. In *ACM MM*, 2003.

[21] L. Marchesotti, C. Cifarelli, and G. Csurka. A framework for visual saliency detection with applications to image thumbnailing. In *CVPR*, 2009.

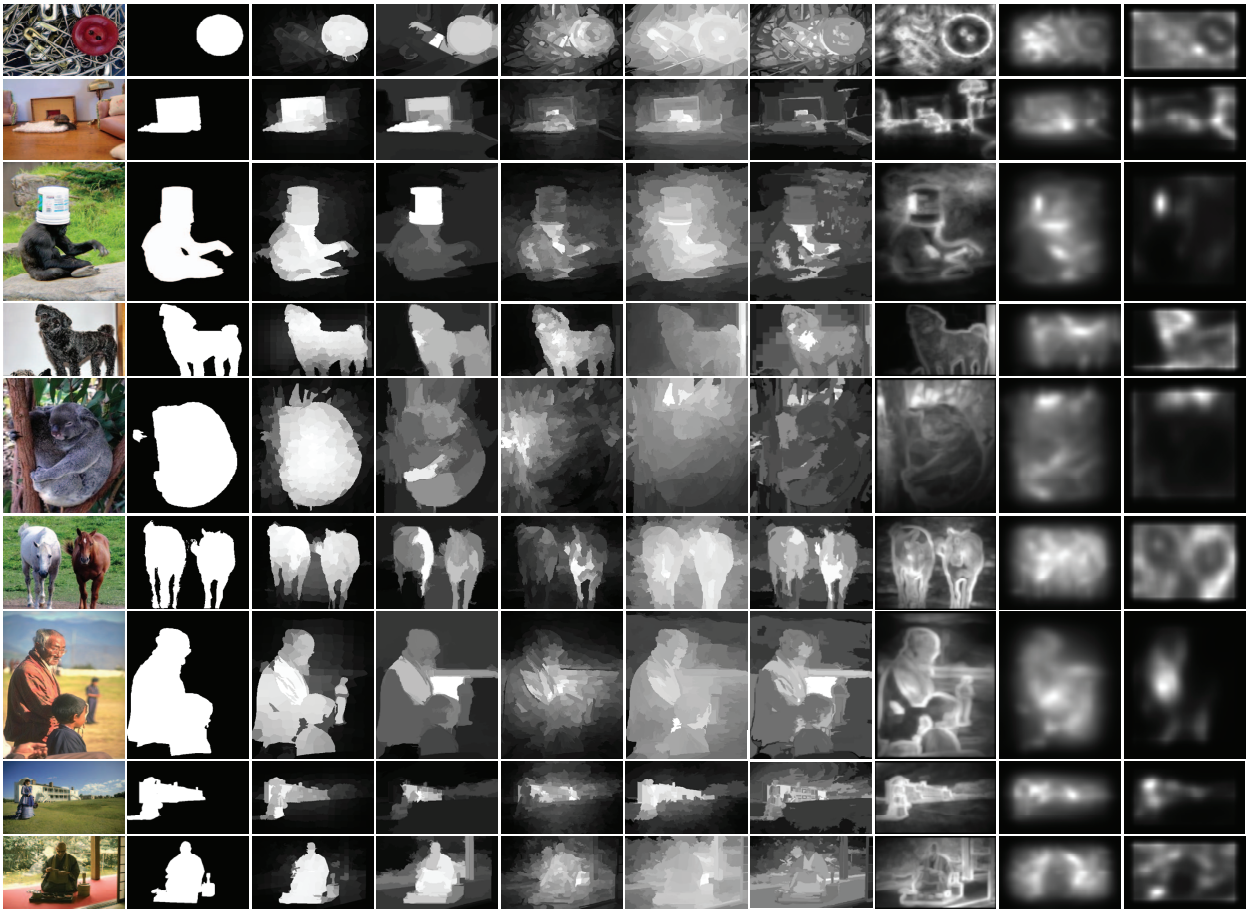

Figure 9. Comparison of different methods on the ASD, SED and SOD datasets. The first three rows are from the ASD dataset, the middle three rows are from the SED dataset, the last three rows are from the SOD dataset.

Table 1. Comparison of average execution time (seconds per image).

Method	Ours	CB	SVO	RC	LR	CA	GB	SER	FT	LC	SR	IT
Time(s)	0.105	1.179	40.33	0.106	11.92	36.05	0.418	25.19	0.016	0.002	0.002	0.165
Code	Matlab	Matlab	Matlab	C++	Matlab	Matlab	Matlab	C++	C++	C++	C++	Matlab

[22] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, 2001.

[23] V. Movahedi and J. H. Elder. Design and perceptual validation of performance measures for salient object segmentation. In *CVPRW*, 2010.

[24] V. Navalpakkam and L. Itti. An integrated model of top-down and bottom-up attention for optimizing detection speed. In *CVPR*, 2006.

[25] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung. Saliency filters: Contrast based filtering for salient region detection. In *CVPR*, 2012.

[26] Z. Ren, Y. Hu, L.-T. Chia, and D. Rajan. Improved saliency detection based on superpixel clustering and saliency propagation. In *ACM*, 2010.

[27] G. S, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. In *CVPR*, 2010.

[28] R. B. S. Alpert, M. Galun and A. Brandt. Image segmentation by probabilistic bottom-up aggregation and cue integration. In *CVPR*, 2007.

[29] X. Shen and Y. Wu. A unified approach to salient object detection via low rank matrix recovery. In *CVPR*, 2012.

[30] L. Wang, J. Xue, N. Zheng, and G. Hua. Automatic salient object extraction with contextual cue. In *ICCV*, 2011.

[31] W. Wang, Y. Wang, Q. Huang, and W. Gao. Measuring visual saliency by site entropy rate. In *CVPR*, 2010.

[32] X.-J. Wang, W.-Y. Ma, and X. Li. Data-driven approach for bridging the cognitive gap in image retrieval. In *ICME*, 2004.

[33] Y. Wei, F. Wen, W. Zhu, and J. Sun. Geodesic saliency using background priors. In *ECCV*, 2012.

[34] C. Yang, L. Zhang, and H. Lu. Graph-regularized saliency detection with convex-hull-based center prior. *IEEE Signal Process. Lett.*, 20(7):637–640, 2013.

[35] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang. Saliency detection via graph-based manifold ranking. In *CVPR*, 2013.

[36] J. Yang and M.-H. Yang. Top-down visual saliency via joint crf and dictionary learning. In *CVPR*, 2012.

[37] Y. Zhai and M. Shah. Visual attention detection in video sequences using spatiotemporal cues. In *ACM MM*, 2006.