

Region saliency detection via multi-feature on absorbing Markov chain

Wenjie Zhang · Qingyu Xiong · Weiren Shi ·
Shuhan Chen

Published online: 1 March 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Saliency region detection plays an important role in image pre-processing, and uniformly emphasizing saliency region is still an intractable problem in computer vision. In this paper, we present a data-driven salient region detection method via multi-feature (included contrast, spatial relationship and background prior, etc.) on absorbing Markov chain, which uses super pixel to extract salient regions, and each super-pixel represents a node. In detail, we first construct function to calculate absorption probability of each node on absorbing Markov chain. Second we utilize image contrast and space relation to model the prior salient map which is provided to foreground salient nodes and then calculate the saliency of nodes based on absorption probability. Third, we also exploit background prior to supply the absorbing nodes and compute the saliency of nodes. Finally, we fuse both the saliency of nodes by cosine similarity measurement method and acquire the ultimate saliency map. Our approach

is simple and efficient and highlights not only a single object but also multiple objects consistently. We test the proposed method on MSRA-B, iCoSeg and SED databases. Experimental results illustrate that the proposed approach presents better robustness and efficiency against the eleven state-of-the-art algorithms.

Keywords Saliency region · Image contrast · Space relation · Background prior · Absorbing Markov chain

1 Introduction

Visual saliency is a concept in neuroscience, psychology, neural systems, and computer vision [1,2]. The main task of saliency detection is to locate the more interested object(s) in a scene and identify them from their neighbors. The extracted saliency map can be served as a pre-processing step for many applications, such as image retrieval [3], image segmentation [4], image retargeting [5], dominant color detection [6], etc.

Two types of saliency detection methods are developed: one [7,8] is top-down and task-driven style and the other [1,2,9–18] is bottom-up and data-driven one. The top-down method focuses on a specific object and learns salient features by supervised learning on a larger data set which contains the specific object, while the bottom-up method relies on some prior knowledge about salient regions and background, such as contrast, compactness, etc. In this paper, we will pay more attention to bottom-up saliency detection.

Bottom-up saliency research has made much breakthrough within the past decades [2,7–20]. In detail, Itti et al. [2] utilize the local center-surround difference to put forward the saliency model based on multi-scale image features. Harel et al. [9] propose a graph-based salient detection method; this method uses edge strengths to denote the dissim-

W. Zhang · W. Shi
College of Automation, Chongqing University,
Chongqing 400044, China
e-mail: daaiyiejian@cqu.edu.cn

W. Shi
e-mail: wrs@cqu.edu.cn

Q. Xiong
Key Laboratory of Dependable Service Computing in Cyber
Physical Society, MOE, Chongqing 400044, China

Q. Xiong (✉)
School of Software Engineering, Chongqing University,
Chongqing 400044, China
e-mail: cquxqy@163.com

S. Chen
College of Information Engineering,
Yangzhou University, Yangzhou, China
e-mail: c.shuhan@gmail.com

ilarity between two nodes on a graph, then regards the most frequently visited nodes as salient nodes in a local context. The main objectives in [9] are to predict human fixations on natural images; the method fails when the background of the scene is cluttered. Hou and Zhang [10] raise a spectral residual approach to detect saliency; the method has a good performance for small salient objects but it has insufficient ability for larger objects since the algorithm regards the large salient object as part of the scene and consequently fails to identify remarkable objects. However, these methods concern more with saliency pixels where the salient object could appear in the image, and the saliency map of these methods are always blurred. Later, many scholars focused more on salient objects or regions in the scene, and salient objects with precise details and high consistency become an important basis for evaluation of the merits of the algorithm. Cheng et al. [11] exploit global contrast differences and spatial coherence to extract salient regions; their method performs well in the case that the salient objects have remarkable contrast features. Achanta et al. [12], compute saliency map using the center-surround principle which compares color traits of each pixel with average values of the whole image. This method is simple and efficient. However, it fails for images with cluttered backgrounds. Achanta and Susstrunk [13] further improve algorithm based on visual difference—maximum symmetric surround saliency (MSSS), which varies the bandwidth of the center surround-filtering. Goferman et al. [14] present a context-aware salient detection algorithm based on four principles of human visual attention. Zhang et al. [15] use region contrast, boundary contrast, smoothness prior, center bias to model a coarse-to-fine saliency and obtain consistent salient object, but this method has limited ability when the salient object is significantly close to the image boundary or when there are background complex scenes. Du and Chen

[1] propose a salient object detection method via random forest which evaluates the saliency based on the rarities of patches and contour-based contrast analysis. Zhai and Shah [16] detect pixel-level saliency through contrast among the pixel to all other pixels, but color information is ignored for efficiency. Chang et al. [17] propose a graphical model to fuse generic objectness and visual saliency together to detect objects, and the results can highlight salient regions; meanwhile, some non-significant regions are reinforced incorrectly in some case. Yang et al. [18] utilize a graph-based manifold ranking algorithm to extract salient objects. Jiang et al. [19] formulate saliency detection via absorbing Markov chain on an image graph model, which bases the boundary prior and sets the virtual boundary nodes as absorbing nodes and the saliency of each node is computed as its absorbed time to absorbing nodes. It performs better in most case. However, the small salient object touching image boundaries may be incorrectly suppressed, and some smooth background regions near the image center are highlighted incorrectly.

Saliency detection has made great progress in recent years, but there are still some issues that remain unresolved. For example, the methods usually require dealing with more background data than the interesting object data, or the methods are inadequate to handle cluttered background, etc. Inspired by Jiang et al.'s [19] method (AMC), we reconsider the prior information including contrast, spatial relationship and background. And then we exploit these image traits to provide the prior saliency information, and utilize the absorbing Markov chain to detect saliency. Our model contains three parts: the first part is the saliency detection via the foreground salient nodes, the second is the saliency detection via background nodes, and the third is an integrated saliency detection method that uses cosine similarity. The main steps of the proposed method are shown in Fig. 1. In detail, our

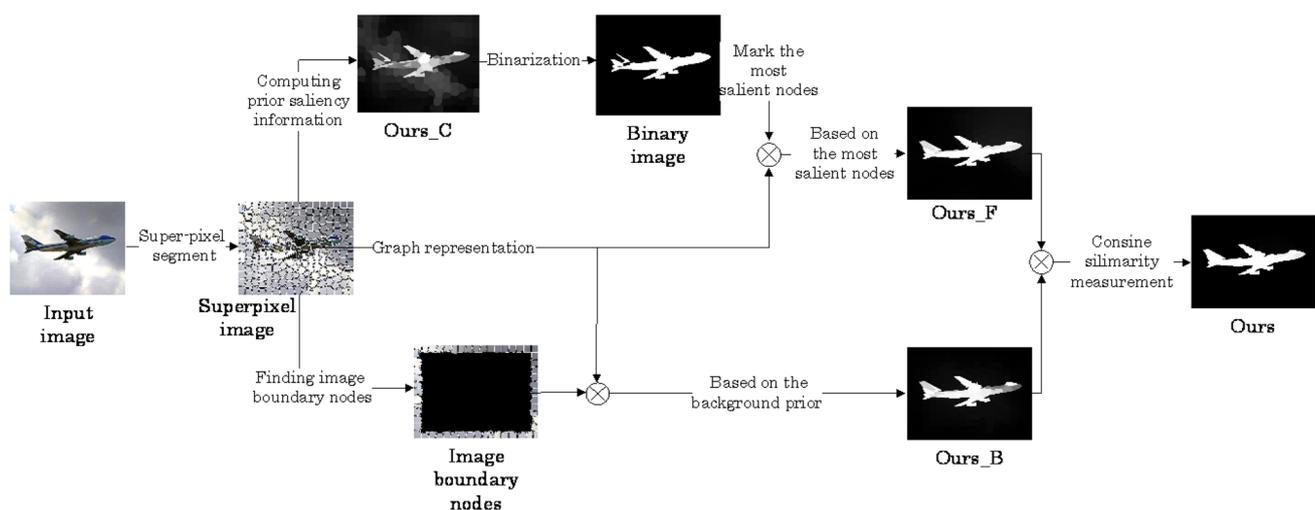


Fig. 1 The flow chart of the proposed method. Note: we regard the Ours_C as method of computing the prior saliency information, the Ours_F as the method based on the foreground salient nodes, and the

Ours_B as method via the background prior. The Ours represents result of the proposed model

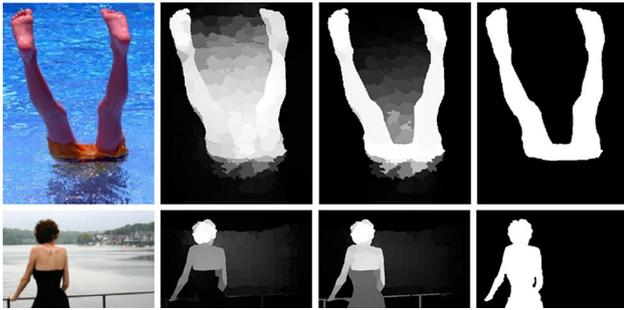


Fig. 2 The results of the proposed method. From *left to right*: the original images, AMC [19], the proposed method, true-ground (salient objects are manually labeled)

approach uses super-pixel method SLIC (simple linear iterative clustering, SLIC [21]) to segment the image into different regions and regards each super-pixel as a node on the graph and then utilizes contrast and space relation to model the prior salient regions. Finally, the proposed method exploits the prior salient region to provide the most salient nodes for absorbing Markov chain by binary segmentation and calculates the absorbing probability of each node by absorbing the Markov chain. Second, we exploit background prior to obtaining the absorbing probability of each node. Finally, we fuse both absorbing probabilities and acquire the final saliency map. We test our method on MSRA-B, iCoSeg and SED databases, and the experimental results indicate that the proposed method can suppress the saliency of non-notable regions near image center as well as image boundary and perform efficiently against the state-of-the-art methods for images with cluttered scene.

Compared with AMC, the main contributions of this work are as follows: (1) We model the prior saliency detection using the images region contrast and spatial distance to provide the prior saliency information, and then detect salient regions based on the foreground salient nodes by absorbing Markov chain, which uniformly strengthens the consistency and coherence of conspicuous regions. (2) We introduce a cosine similarity measurement method and model an integrated saliency maps, which achieves favorable results. If there are long-range smooth background regions near the image center, it is intractable issue to use the absorbed time to obtain the salient regions. In AMC algorithm, AMC exploits the equilibrium probability to regulate the absorbed time so as to suppress the saliency of this kind of regions. However, it is not always effective. In this paper, we combine the absorbing probability based on the most salient nodes and the absorbing probability-based background prior during saliency detection and try to solve this issue. The examples of AMC and the proposed method are shown in Fig. 2.

The remainder of this paper is organized as follows: In Sect. 2, we introduce absorbing Markov chain fundamentals and construct function to calculate absorption probability

of each node. Section 3 details the process that created the graphs and presents the analysis of absorbing Markov chain on k -regular graph. And in Sect. 4, we propose our saliency detection approach. Experimental results and analyses are given in Sect. 5, and conclusion is shown in Sect. 6.

2 Absorbing Markov chain fundamentals

The absorbing Markov chain is a semi-supervised learning algorithm. By marking a set of given nodes, this paper regards these labeled nodes as absorbing nodes and the remaining nodes as transient nodes. Then the absorbing probabilities which random walker moves from transient nodes to absorbing node can be obtained by absorbing Markov chain, and so the absorbing probability reflects the relationship between absorbing node and transient node. The goal is to learn absorbing probability moving from transient nodes to absorbing node. During the saliency detection, conspicuous regions always have the similarity. Therefore, we utilize the absorbing probability to represent the saliency of nodes.

This section succinctly states some fundamental results of absorbing Markov chain [22–24] and then calculates the probability of moving from each node to the absorbing node.

Let $S = \{s_1, s_2, s_3, \dots, s_m\}$ be a set of states (or nodes), a Markov chain can be completely specified by the $m \times m$ transition matrix P , where p_{ij} is the probability of moving from state s_i to state s_j . On absorbing Markov chain, random walker starting at any transient state reaches absorbing state and cannot leave from the absorbing state (not just in one step), which indicates that any pair of absorbing nodes are unconnected. To assume that an arbitrary absorbing Markov chain has r absorbing states and t transient states, and renumbering the states making the transient states comes first, the transition matrix P will have canonical form as follows:

$$P = \begin{pmatrix} Q & R \\ \mathbf{0} & I \end{pmatrix}, \quad (1)$$

where Q is a t -by- t matrix which contains transient probability between any two transient states, while R is a nonzero t -by- r matrix and contains the probabilities moving from each transient state to each absorbing state, $\mathbf{0}$ is an r -by- t zero matrix and I is an identity matrix.

For the transition matrix P , the fundamental matrix $N = (I - Q)^{-1}$ can be derived from P . The entry n_{ij} of N can be described as the probability that random walker starts from the transient state s_i to the transient state s_j . Let b_{ik} be the probability that the transient state s_i be absorbed in absorbing state s_k , and B is the matrix with entries b_{ik} . Then B is computed as

$$B = NR, \quad (2)$$

where B is a t -by- r matrix and R is as in the canonical form.

The i th row of B represents the absorption probabilities starting from the transient state s_i to each absorbing state. If random walker starting from the transient state s_i arrives at the absorbing state s_k with larger probability, the saliency of transient node s_i will be closer to absorbing node s_k since the saliency of absorbing node s_k is known. Therefore, the proposed method uses the following formula (3) to calculate the final absorption probability of each node. We verify the proposed method in Sect. 5:

$$f_i = \begin{cases} 1, & s_i \in R_1(s_i) \\ \max_{1 \leq k \leq r} \{b_{ik}\}, & \text{otherwise} \end{cases}, \tag{3}$$

where f_i denotes the probability which node s_i is absorbed. $R_1(s_i)$ is the labeled node set.

3 Graph representations

Given an input image represented as an absorbing Markov chain, the probability matrix P can be constructed by a single-layer graph $G(V, E)$, where V is the set of states or nodes, and E is the set of edges. In this work, each node is a super pixel generated by the SLIC algorithm [21]. Since neighboring nodes may possess similar appearance and notability, the edges can be represented though the k -regular graph. On the k -regular graph, each node is connected to the nodes which neighbor it or share common boundaries with its neighboring nodes. The edge weights between nodes can be expressed by affinity matrix W , in which high weight is regarded as strongly connected pair of nodes, and low weights denote nearly disconnected nodes. With the constraints on edges, the k -regular graph is a sparsely connected. i.e., most elements of the affinity matrix W are zero. In this work, the weight w_{ij} between two nodes is expressed by

$$w_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|}{\sigma^2}}, & j \in R_2(i), \\ 0, & \text{otherwise} \end{cases}, \tag{4}$$

where super-pixel x_i and x_j are denoted by the mean of the pixels in corresponding super-pixel image region in the CIE LAB color space. The super pixel nodes are normalized to the range [0 1] through the maximum. The constant σ is fixed to control the strength of the weight. $R_2(i)$ is the neighboring node sets of x_i .

In order to calculate the probability transition matrix P , we define a new affinity matrix A to signify the relation of nodes as (5). The row weights of the affinity matrix A need to be divided by the degree of the corresponding nodes to get the probability transition matrix. In this paper, we define the diagonal matrix D to normalize the row of A , $D = \text{diag}\{\sum w_{1j}, \sum w_{2j}, \dots, \sum w_{rj}\}$, Finally, the transient matrix P is given as (6)

$$a_{ij} = w_{ij} \times \text{sign}(w_{ij}) \tag{5}$$

$$P = D^{-1} \times A \tag{6}$$

where $\text{sign}(w_{ij})$ is a symbolic function, $\text{sign}(w_{ij}) = 1$ if the node x_i is a transient node or $i = j$, else $\text{sign}(w_{ij}) = 0$.

In this way, the random walker is restricted to a local region while its path is determined by the k -regular graph. Absorption probability moving from transient node to absorbing node is affected by spatial distance and transition probabilities. i.e., the node will obtain greater absorption probability if it has larger transition probability and is closer to the absorbing node.

4 Saliency detection model

Assuming an input image represented as a graph, the following task is to identify the absorbing nodes that most likely belong to the salient regions or background regions in the image. In this paper, we calculate absorption probabilities which move from each transient node to salient region, and absorption probabilities moving from each transient node to the background region, respectively. Then we calculate the integrated saliency map by a cosine similarity measurement method. The following subsection will describe the process of the proposed method.

4.1 Saliency detection via the foreground salient nodes

In this subsection, we introduce how to discover the salient nodes based on image contrast and spatial distance information and mark the most significant nodes as the absorbing nodes by the binary segmentation of Otsu [25] method. Otsu takes the maximum variance between foreground regions and background regions as threshold selection criteria and achieves better segmentation results. And the threshold is calculated as (7)

$$\delta^2(T_A) = \max_{0 \leq T \leq L} \delta^2(T), \tag{7}$$

where $\delta^2(\cdot)$ is the variance between salient regions and non-salient regions, T denotes the threshold, and L is the maximum of pixels. T_A represents the threshold if the variance takes the maximum.

In visual attention process, those unique, unpredictable, scarcity and the singularity of the object is to draw attention, and other objects or background are of less concern. Image contrast and spatial relationship are important features for image saliency in previous saliency research [11, 14, 26]. In general, people pay more attention to the image regions that contrast strongly with the neighboring regions. Besides, high contrast to its surrounding regions usually easily attracts

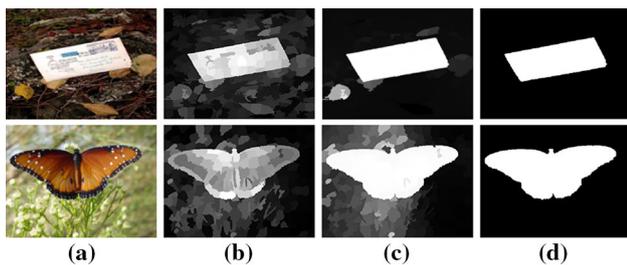


Fig. 3 The saliency map based on the most salient nodes. From left to right: **a** the original images, **b** Ours_C, **c** Ours_F (saliency detection based on foreground salient nodes), **d** true-ground

attention than high contrast to far-away regions. In addition, ‘center prior’ is considered in some previous saliency models [27]. For center prior, the nodes are more salient if these nodes are closer to the image center. It is valid in many cases. However, it is not always effective in general cases. In our work, we utilize these prior visual saliency information to model a salient region detection. In the process of detecting saliency, we take ‘center prior’ as a smaller weight factor to avoid over-enhancing insignificant regions near image center. The significant contribution degree $\Psi(x_i)$ for the super pixel node x_i can be calculate as

$$\Psi(x_i) = \frac{1}{1 + c \cdot d_c(x_i)} \sum_{j=1}^K \frac{\|x_i - x_j\|}{1 + \alpha \cdot d_p(x_i, x_j)}, \tag{8}$$

where $c(0 < c < 1)$ is the ‘center prior’ weight parameter, α is spatial distance parameter, K is the total number of super-pixels. The $d_c(x_i)$ is the Euclidean distance from the super-pixels x_i to the image center and normalized to the range [0 1]. This paper regards the centroid of the super pixel region as super-pixel spatial position. The $d_p(x_i, x_j)$ is Euclidean distance between super pixel x_i and x_j .

The super-pixel x_i is salient when $\Psi(x_i)$ is high. Hence, the prior saliency of the node x_i can be calculated as

$$S_{\text{priori}}(x_i) = 1 - e^{-\Psi(x_i)}, \tag{9}$$

where $S_{\text{priori}}(x_i)$ is the prior saliency map.

We represent the method of (9) as Ours_C; the result can be seen in Fig. 3b. Although this proposed approach has limited capacity to highlight consistency of the significant object or regions, the prior saliency map can provide effective saliency information.

We reconsider the absorbing Markov chain model to improve the consistency of saliency detection. In detail, we mark the most salient nodes as absorbing nodes by binarizing $S_{\text{priori}}(x_i)$. The threshold is selected by (7) so that salient nodes are labeled as accurately as possible. And we regard the node x_i as the most salient nodes if $S_{\text{priori}}(x_i) > T_A$. In detail, we label these salient nodes as absorbing nodes, and

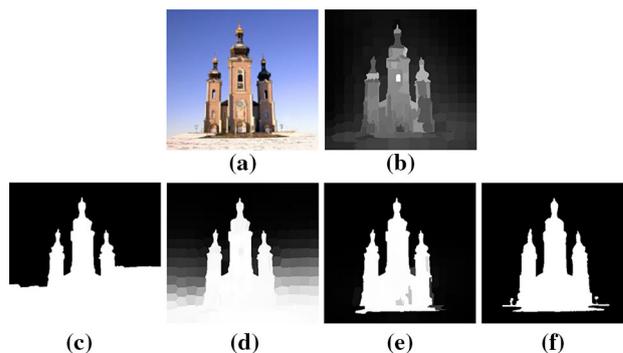


Fig. 4 The failure example based on most salient nodes. From left to right: **a** the original images, **b** Ours_C, **c** the binary image, **d** Ours_F, **e** Ours, **f** true-ground

the remaining nodes as transient nodes. Then we can get the transition matrix P and calculate absorption probability of the node x_i by (3). The absorbing nodes belong to salient nodes, so super-pixel x_i is salient if random walker starting from x_i to absorbing nodes has a large absorption probability. Therefore, the saliency map $M_f(x_i)$ based on the most salient nodes can be represent as

$$M_f(x_i) = f_i \tag{10}$$

This method based on foreground salient nodes can be regarded as Ours_F. It can improve the consistency of the salient object, as seen in Fig. 3c and this method is valid in most case. However, when the contrast of background regions are high in some case, and consequently some background nodes are labeled as absorbing states incorrectly by binary segmentation (see Fig. 4c), it leads to some background regions enhanced as well as salient objects (see Fig. 4d). To alleviate this problem and further improve the performance, we utilize the boundary prior to inhibit the saliency of non-salient nodes. The following subsection gives detailed explanation.

4.2 Saliency detection via background prior

The background often manifests local or global appearance connectivity with each of four image boundaries as salient objects less likely occupy all four image boundaries [18, 19, 28] and background regions often connect with image boundaries. Inspired by this prior saliency information, we describe the image boundaries’ nodes as the absorbing nodes; therefore, the random walker starting in background nodes will arrive at the absorbing nodes with larger absorbing probability. That is, larger absorption probability will indicate lower saliency for the nodes. So the saliency map $M_b(x_i)$ via the background nodes can be denoted as (11). Specifically, the transition matrix P based on boundary prior can be got by

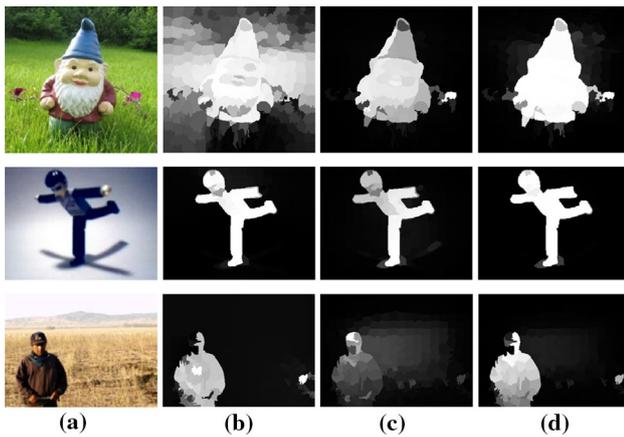


Fig. 5 The results of background-based saliency map. **a** The original images, **b** Ours_F, **c** Ours_B, **d** Ours (the proposed method)

(6), we can easily extract matrix Q and R by (1), and the fundamental matrix N is calculated relied on Q , and absorption probability matrix B is computed by (2). Finally, we obtain the absorption probability of nodes by (3).

$$M_b(x_i) = 1 - f_i \quad (11)$$

Figure 5c shows results of proposed approach (Ours_B) in this subsection. The saliency map $M_b(x_i)$ can suppress the non-salient regions better and protrude remarkable regions, but it is noted that the Ours_B has poor performance to detect salient object when the object touches image boundaries, as shown in third saliency map of Fig. 5c, while the presented model (Ours_F) in Sect. 4.1 can avoid this issue effectively. The Ours_F can enhance the uniformity of salient object, and it does not matter whether the object close to image boundary. And the Ours_B can suppress the background better than Ours_F when some background regions have high contrast. The saliency measures by Ours_F and Ours_B are complementary to each other.

4.3 Cosine similarity measurement of saliency maps

In this paper, we integrate Ours_F method with Ours_B method to improve the performance by cosine similarity measurement. The node x_i always is salient if both $M_f(x_i)$ and $M_b(x_i)$ are large. We introduce the similarity measurement to evaluate similarity of both methods. $M_f(x_i)$ and $M_b(x_i)$ are larger, illustrating that they are more similar and the node x_i is more likely to be a significant node. Thereby we compute integrated saliency map relied on similarity measurement. Similarity measurement estimates the difference between two individuals [29,30]. In this work, we evaluate similarity between $M_f(x_i)$ and $M_b(x_i)$ using extended cosine similarity function $\text{Sim}_g(M_f(x_i), M_b(x_i))$, which is defined as

$$\begin{aligned} & \text{Sim}_g(M_f(x_i), M_b(x_i)) \\ &= \frac{M_f(x_i) \times M_b(x_i)}{\|M_f(x_i)\|^2 + \|M_b(x_i)\|^2 - M_f(x_i) \times M_b(x_i)}, \quad (12) \end{aligned}$$

where $\text{Sim}_g(M_f(x_i), M_b(x_i)) \in [0,1]$, the $\text{Sim}_g(M_f(x_i), M_b(x_i))$ closer to 1 indicates smaller difference between $M_f(x_i)$ and $M_b(x_i)$.

The node x_i has higher saliency when both $M_f(s_i)$ and $M_b(s_i)$ are closer to 1, so we calculate an integrated saliency $S(M_f(s_i), M_b(s_i))$ based on extended cosine similarity by (13).

$$\begin{aligned} & S(M_f(s_i), M_b(s_i)) \\ &= \text{Sim}_g(M_f(s_i), M_b(s_i)) \times \frac{M_f(s_i) + M_b(s_i)}{2} \quad (13) \end{aligned}$$

The examples of final results are shown in Fig. 5d. It is worth noting that cosine similarity measurement enforces these two maps to serve as the prior and cooperate with each other in an effective manner, which suppresses the background and uniformly highlights the salient regions in an image.

5 Experiments

To validate our proposed approach, we evaluate our model in terms of precision, recall, F_β , mean absolute error (MAE) and precision-recall curve (PR curve). At the same time, we compare our method against state-of-the-art algorithms (IT [2], GBVS [9], SR [10], RC [11], FT [12], RA [13], CA [14], LC [16], SVO [17], GBMR [18] and AMC [19]). Most of these algorithm codes are available in the authors' home page. Our experiments are performed on three datasets: MSRA-B, iCoSeg and SED.

5.1 Data sets of experiment

The MSRA-B [7] contains 5,000 images, and salient objects were manually labeled by Jiang et al. [31]. MSRA-1000 is a subset of the MSRA-B with 1,000 images. And the iCoSeg [32,33] is a co-segmentation set, and provides 38 groups of 634 images, along with pixel ground-truth hand annotations, and we use it to evaluate the performance of detecting saliency. The SED [34] has two subsets: one is SED1, which has 100 images, and each image contains a significant object; the other is SED2 with 100 images and each image has two significant objects. The SED also provides annotation with the labeled salient object for each image.

5.2 Evaluation metrics

For each method, the precision and recall for an image are calculated by segmenting each saliency map into a binary

map with a given threshold $T_1 \in [0, 255]$ and then comparing with the ground truth mask. The precision value is the ratio of salient pixels correctly assigned to all the pixels of extracted regions, which reflects the accuracy of the detection algorithm. The recall value corresponds to the percentage of detected salient pixels in relation to the ground-truth number, which represents the detection consistency. The precisions and recall can be depicted by the PR curve on the data set. The precision and recall rate for each image are quantified as follows:

$$\text{Precision} = \frac{\sum_{i=1}^W \sum_{j=1}^H B(i, j) \cdot G(i, j)}{\sum_{i=1}^W \sum_{j=1}^H B(i, j)} \tag{14}$$

$$\text{Recall} = \frac{\sum_{i=1}^W \sum_{j=1}^H B(i, j) \cdot G(i, j)}{\sum_{i=1}^W \sum_{j=1}^H G(i, j)}, \tag{15}$$

where B is the binary salient object mask generated by thresholding saliency map and G is the corresponding binary ground truth. W and H are the width and height of the saliency map.

The F_β is a weighted harmonic mean between the precision and recall values, which is the overall performance measurement. Different from calculating PR curve, we exploit the fixed and adaptive thresholding TH in the process of generating binary salient object masks. F_β is defined as (17).

$$\text{TH} = \frac{2}{W \times H} \sum_{i=1}^W \sum_{j=1}^H S_{\text{map}}(i, j) \tag{16}$$

$$F_\beta = \frac{(1 + \beta^2) \times \text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}}, \tag{17}$$

where $\beta^2 = 0.3$ stresses precision more than recall, similarly to [11, 12].

The MAE is a statistical measure that represents the difference between estimates and actual values. In this paper, the MAE is utilized to estimate the dissimilarity between the saliency map and ground truth. And the lower MAE value indicates better performance. The MAE is the average of absolute error between the continuous saliency map S_{map} and the binary ground truth G , which is defined as

$$\text{MAE} = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H |S_{\text{map}}(i, j) - G(i, j)| \tag{18}$$

5.3 Performance comparison

Experimental setup. For presented approach, we set the number of super-pixels $K = 250$ and discuss the effects of changes of super-pixel number K value on the proposed method in Exp. 1. In Eq.(4), the weight σ^2 is set to con-

trol the strength of weight between a pair of nodes $\sigma^2 = 0.1$, using the same setting as [15, 18, 19]. The ‘center prior’ parameter c to weight the impact of ‘center prior’ and spatial distance parameter α is used to control influence of spatial distance in Eq. (8); we take $c = 0.2$ and $\alpha = 0.7$. All experiments are tested on a Dual Core 2.8 GHz machine with 2 GB RAM.

Exp. 1: the effects of changes of super-pixel number K on the proposed approach. In this paper, the presented schedule utilizes super pixel method SLIC [21] to pre-process image and then detects distinctive regions. The paper assesses the impact of super pixel number K on the proposed method, and quantitative results comparison has been made by setting different supper pixel number K to guide the selection of K ; the PR curves on the iCoSeg are shown in Fig. 6. In detail, Fig. 6a gives the PR curves of Ours_C for different K , and Fig. 6b shows the PR curves of the final result of the proposed algorithm (Ours) for different K .

As shown in Fig. 6a, when K changes from 50 to 250, the PR curves of Ours_C can be improved. While PR curves’ performance of Ours_C is similar between $K = 250$ and $K = 300$. Meanwhile, the PR curves of Ours perform better when K equals 250 or 300 (see Fig. 6b). And the average running time of proposed method is given in Table 1; it can be found that the proposed method has the longer average run time for larger K . Therefore, considering the computational complexity and the performance of PR curves, we select supper number $K = 250$ for all experiments.

Exp. 2: comparisons of the three parts for proposed method. In this experiment, we evaluate our method based on prior saliency information (Ours_C) and the results of the proposed method (including Ours_F, Ours_B, Ours) in terms of F_β , precision and recall. The results on MSRA-1000 can be seen in Fig. 7. Inspired by AMC [19], we also compare the result of AMC with the proposed method. The AMC regards the saliency of nodes as the expected time, which the nodes start from the transient state and arrive at the absorbing state on the absorbing Markov chain.

Figure 7 shows the average precision, recall and F_β . Compared with the Ours_B (saliency detection via background prior), the Ours_F (saliency detection via the foreground salient nodes) has better performance in terms of recall, but the Ours_F strengthens the non-significant regions in some cases, which causes lower precision and F_β . On the other hand, the Ours_B can inhibit the background, and has higher precision and F_β against the Ours_F. The proposed method (Ours) integrates the Ours_F and the Ours_B; although its precision scores are 1.5% lower than the Ours_B, its recall and F_β perform better. In addition, we compare our method with the AMC; our algorithm improves effectively.

Exp. 3: the sensitivity of the proposed method to noise. The Salt and Pepper noise and Gaussian White noise are

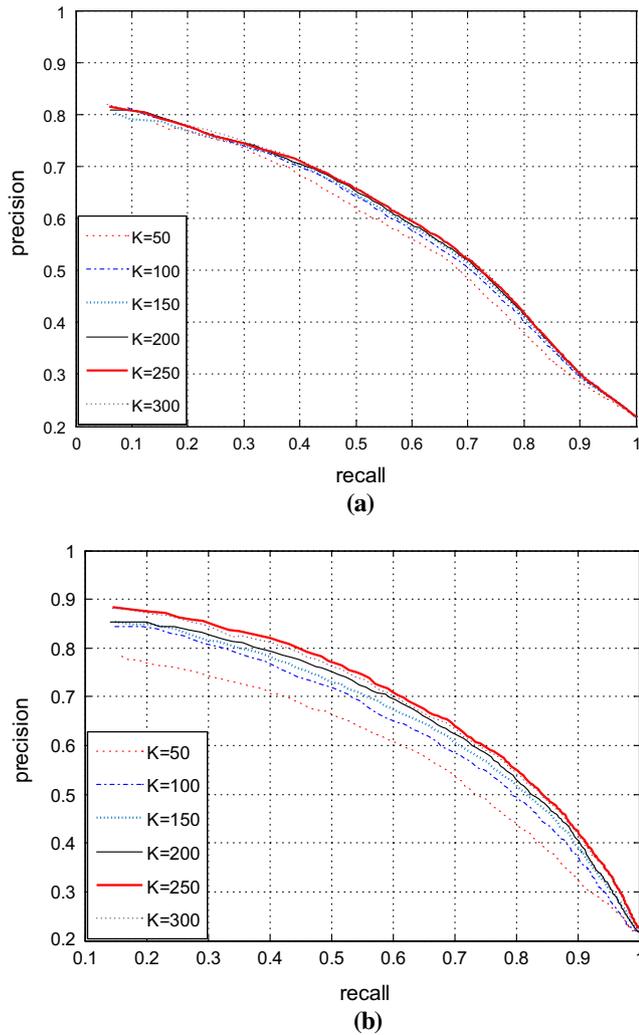


Fig. 6 The effects of changes of super-pixel number K on the proposed approach. **a** The PR curves of Ours_C by setting $K = 50, K = 100, K = 150, K = 250, K = 300$ on iCoSeg database. **b** The PR curves of Ours by setting $K = 50, K = 100, K = 150, K = 250, K = 300$ on iCoSeg database

Table 1 Average running time by setting different super pixel number K in the iCoSeg database

K	50	100	150	200	250	300
Time (s)	0.52	0.57	0.66	0.73	0.81	0.90

The mean size of image in iCoSeg is 506.8×532.7

employed to measure the sensitivity of the proposed method to noise. Two group saliency maps of noise images have been shown in Fig. 8, and the quantitative results have been given in Fig. 9.

In detail, the experiment sets the Salt and Pepper noise density varying from 0.01 to 0.35, and tests the effects of Salt and Pepper on the algorithm; the visual results can be seen in Fig. 8a, and the weighted harmonic mean F_β of pro-

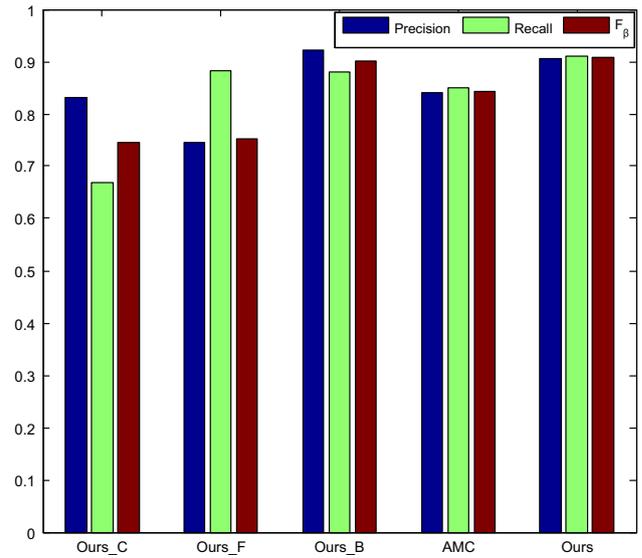


Fig. 7 The comparisons of the three parts for proposed method on MSRA-1000

posed method is shown in Fig. 9. We regard OursNSP as the relationship between F_β and Salt and Pepper noise density. Simultaneously, this paper also utilizes images containing the Gaussian White noise to assess the proposed approach. The variance of Gaussian White noise is regarded as noise density varying from 0.01 to 0.35, and the mean is zero. The detection result of Gaussian White noise images can be seen in Fig. 8b. The F_β -Gaussian-White-noise curve is represented by OursNGW.

As illustrated in Fig. 9, the proposed algorithm can better suppress the influence of the Salt and Pepper noise than the influence of the Gaussian White noise. For the Salt and Pepper noise, when the noise density is less than 0.15, the weighted harmonic means F_β is higher than 0.6. The F_β can retain higher than 60% if only the noise density of Gaussian White noise is less than 0.03. Therefore, the proposed method has better robustness when noise density is less 0.03. It is worth noting that the presented method can also suppress Salt and Pepper noise well if the noise density is less than 0.15.

Exp. 4: quantitative comparison of the MAE. The MAE is utilized to evaluate the proposed approach against the 11 state-of-the-art methods on MSRA_B; the results can be seen in Fig. 10. It is weaker for SVO algorithm to inhibit the non-salient region, and it consequently leads to the larger MAE. AMC and GBMR highlight the prominent regions and therefore they have smaller MAE. Compared with the GBMR, the result of the proposed algorithm is lower, which indicates that our method has higher consistency in terms of MAE.

Exp. 5: quantitative comparison of PR curves. The PR curves of the 11 algorithms mentioned on three databases are provided in Fig. 11. The MAC, GBMR and the pro-



Fig. 8 The sensitivity of the proposed method to noise. **a** The Salt and Pepper noise images and their saliency map. From left to right: the noise density is 0.01, 0.05, 0.1, and 0.15 in Salt and Pepper noise images. **b**

The Gaussian White noise images and their saliency map. From left to right: the noise density is 0.01, 0.02, 0.03, and 0.05 in Gaussian White noise images

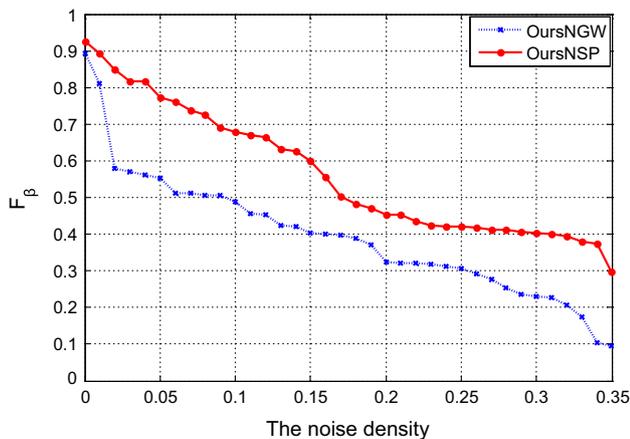


Fig. 9 The effects of noise on the proposed method

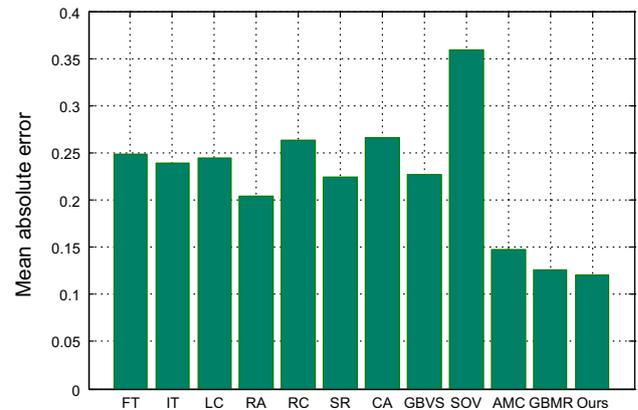


Fig. 10 The MAE results of the proposed method and eleven the state-of-the-art methods on MSRA_B

posed method have better performance than the other methods on the MSRA-B and SED1 datasets, as shown in Fig. 11a, c. This illustrates that the presented method is desir-

able for detecting single significant object since the image always has one object on MSRA-B and SED1 datasets. The proposed algorithm has better performance than the other

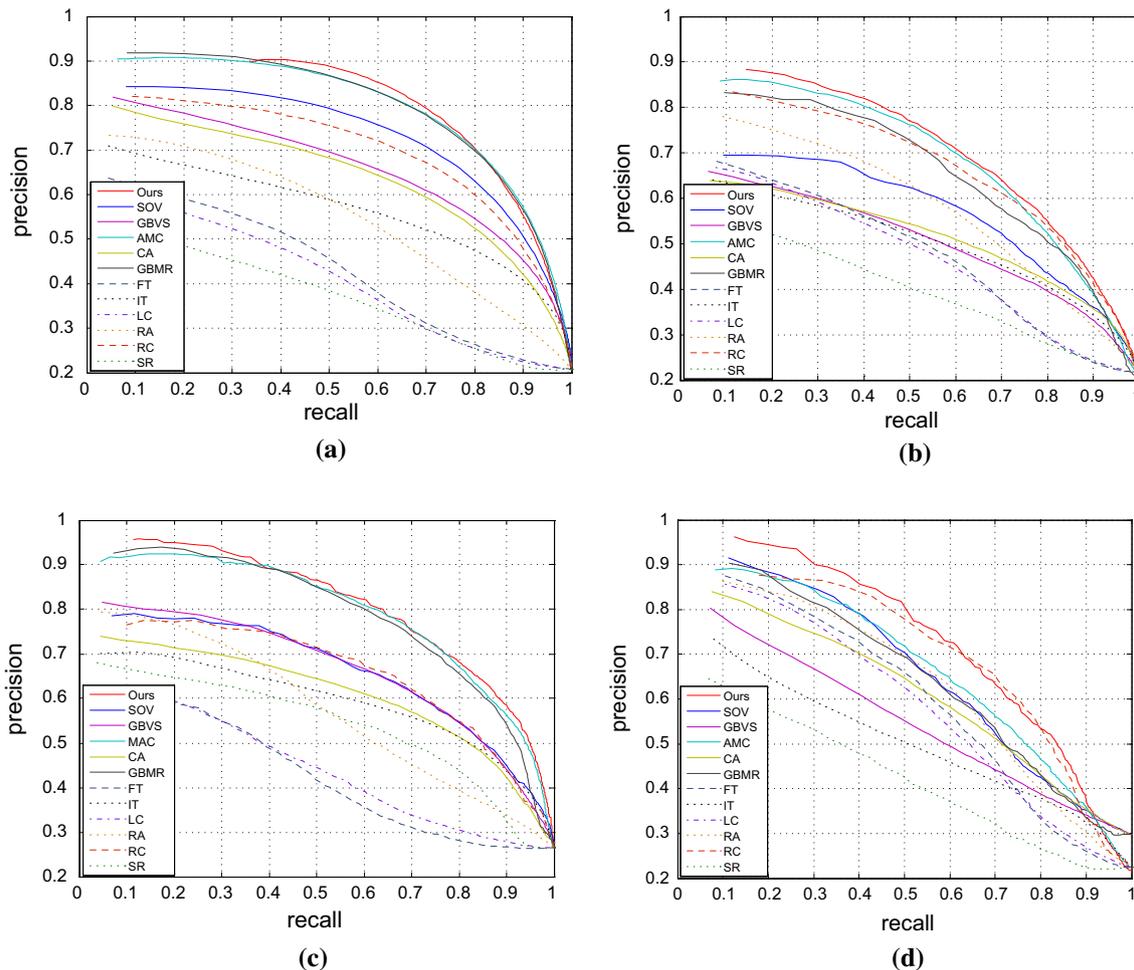


Fig. 11 Quantitative comparison of saliency methods on three image databases. **a** The MSRA-B database, **b** the iCoSeg database, **c** the SED1 database, **d** the SED2 database

methods on the ICoSeg and SED2 datasets, which is shown on Fig. 10b, d. There are one or more remarkable objects in an image on the ICoSeg datasets; therefore, our algorithm is robust for multi-object scene. In general, the presented method is satisfactory in terms of PR curve on three databases.

Exp. 6: qualitative comparison. We provide the visual comparison of different methods in Figs. 12, 13, and 14. The true grounds are provided at the same time. The GBMR, AMC, and the proposed methods belong to semisupervised learning algorithm. Since GBMR and AMC show over-reliance on background priori, it results in nonsignificant regions around the center being enhanced or salient regions touching image boundaries being suppressed incorrectly in some cases; the second saliency maps in Fig. 13e, f are the fail examples. The proposed approach utilizes regional contrast, spatial relationship to detect remarkable region and suppresses non-salient region near image center or image boundaries, as shown in second saliency map of Figs. 13g and 14g. The RC method has obvious advantages when large contrast

differences exist between salient object and background, as shown in first saliency map of Fig. 13d, but the contrast is not always effective in some cluttered background, as shown in the former two saliency maps in Fig. 12d. Our model evaluates image saliency by cosine similarity measurement; the results of the proposed method can highlight salient regions better than other methods in messy sense (see Fig. 12e). The GBVS method forces on salient points and the prominence objects are imprecision in saliency maps. In summary, the proposed method is effective to strengthen the consistency of salient object, and our method performs well for cluttered sense.

5.4 Running time

Table 2 shows the average time taken by each method for all the 5,000 images in the MSRA-B database. Compared with IT, FT, RA, SR, CA, GBMR and AMC, the proposed approach has longer execution time. But our approach performs better in terms of PR curves and the MAE. Note: all

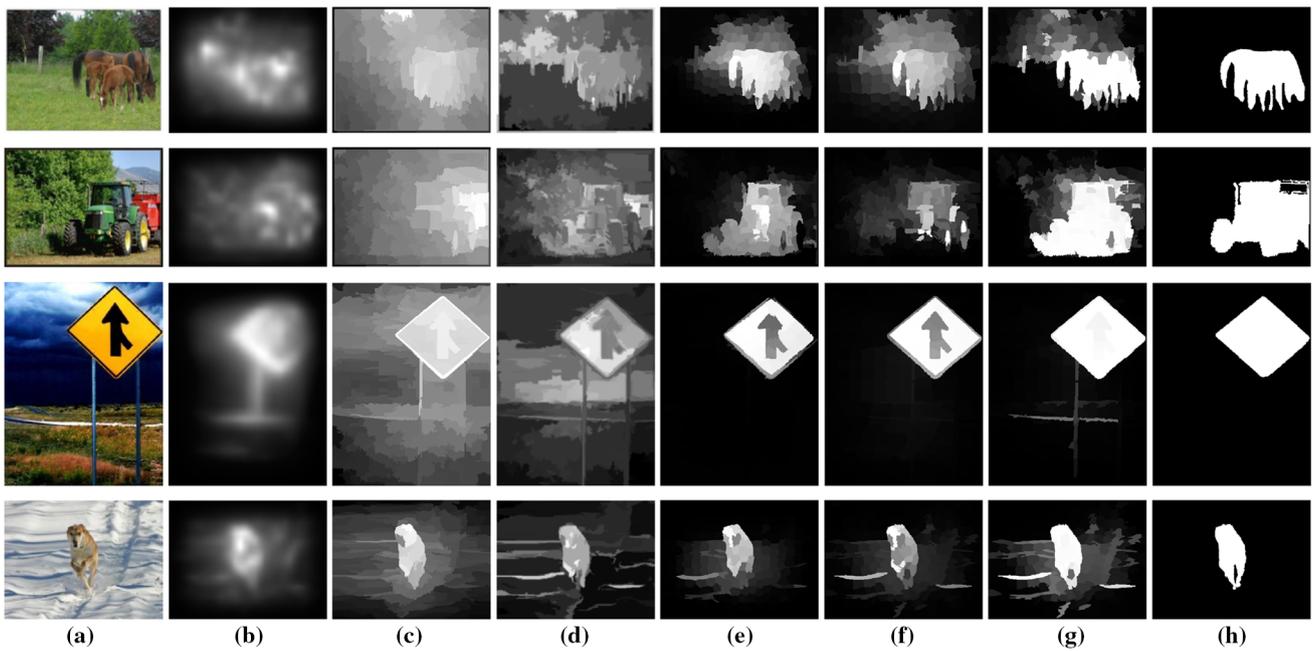


Fig. 12 The saliency maps of different methods on the MSRA-B database. **a** The original images, **b** GBVS, **c** SVO, **d** RC, **e** GBMR, **f** AMC, **g** Ours, **h** true-ground

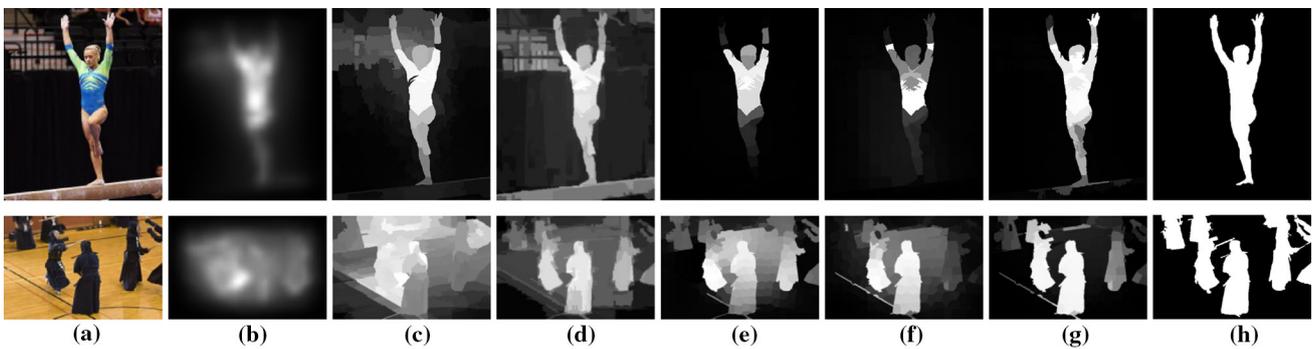


Fig. 13 The saliency maps of different methods on the iCoSeg database. **a** The original images, **b** GBVS, **c** SVO, **d** RC, **e** GBMR, **f** AMC, **g** Ours, **h** true-ground

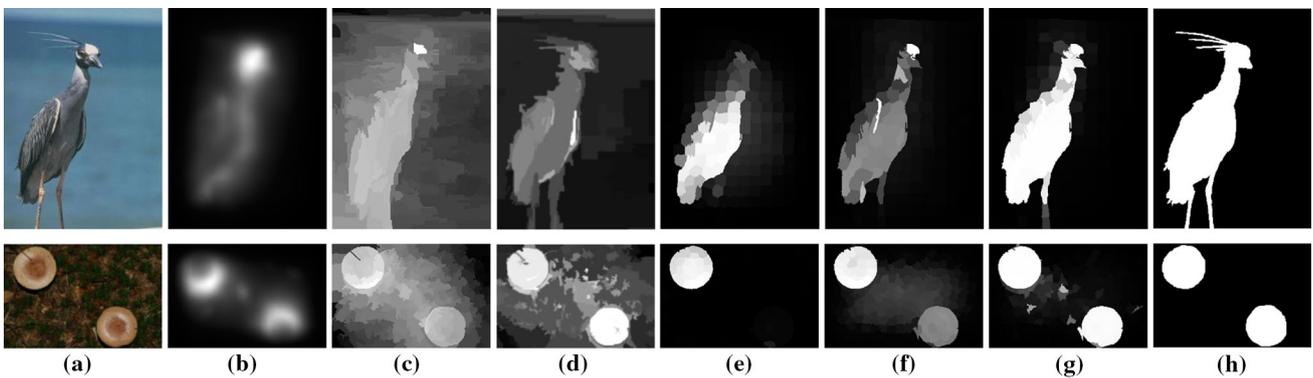


Fig. 14 The saliency maps of different methods on the SED database. **a** The original images, **b** GBVS, **c** SVO, **d** RC, **e** GBMR, **f** AMC, **g** Ours, **h** true-ground. From *top* to *bottom*: the SED1 database, the SED2 database

Table 2 Average running time taken to compute a saliency map for images in the MSRA-B database

Method	IT	FT	RA	SR	CA	SVO	GBVS	GBMR	AMC	Ours
Time (s)	0.25	0.266	0.54	0.07	72.29	53.42	2.085	0.255	0.193	0.654

The mean size of image in MSRA-B are 293.2×399.6

the compared algorithms are implemented in matlab so as to enhance the comparability of the different algorithms. The super pixel generation by SLIC [21] spends 0.163 s, we did not consider the running time of SLIC in GBMR, AMC and the proposed method.

6 Conclusions

We incorporated regional contrast, spatial relationship, center prior and background prior to extract salient regions on absorbing Markov chain. The proposed method detected salient regions on super-pixel image, which made our method process less image data. The saliency detection based on the foreground salient nodes (Ours_F) was proposed, which strengthens the consistency and coherence of noteworthy regions. And the saliency detection via background prior (Ours_B) highlighted the notable regions. Finally, we introduced an integration method by cosine similarity measurement, which makes detection result perform better than Ours_F and Ours_B in terms of recall and F_β . Experimental results on three databases show that the proposed method suppresses the non-salient regions and consistently outperformed existing saliency detection methods on cluttered sense, yielding a satisfactory PR curve as well as visual quality. Meanwhile, the presented approach can suppress the Salt and Pepper noise and Gaussian White noise well when noise density is less than 0.03. In future work, we will optimize running time or build a new model by incorporating high-level knowledge, which makes the algorithm have even better performance, and consider sensitivity of the method to higher density noise.

Acknowledgments The authors thank the anonymous reviewers for helping to review this paper. This work was supported by Major State Basic Research Development Program (973 Program Grant no. 2013CB328903), special fund of 2011 Internet of Things development of Ministry of Industry and Information Technology (2011BAJ03B13-2) and Chongqing Key Project of Science and Technology of China (cstc2012gg-yyjs40008).

References

- Du, S., Chen, S.: Salient object detection via random forest. *IEEE Signal Process. Lett.* **21**, 51–54 (2014)
- Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(11), 1254–1259 (1998)
- Wang, X., Ma, W., Li, X.: Data-driven approach for bridging the cognitive gap in image retrieval. In: *IEEE International Conference on Multimedia and Expo*, pp. 2231–2234 (2004)
- Ko, B.C., Nam, J.Y.: Object-of-interest image segmentation based on human attention and semantic region clustering. *J. Opt. Soc. Am. A (JOSA A)* **23**(10), 2462–2470 (2006)
- Wang, D., Li, G., Jia, W., Luo, X.: Saliency-driven scaling optimization for image retargeting. *Vis. Comput.* **27**(9), 853–860 (2011)
- Wang, P., Zhang, D., Wang, J., Wu, Z., Hua, X. S., Li, S.: Color filter for image search. In: *Proceedings of the 20th ACM International Conference on Multimedia*, pp. 1327–1328 (2012)
- Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., Shum, H.Y.: Learning to detect a salient object. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(2), 353–367 (2011)
- Yang, J., Yang M.: Top-down visual saliency via joint crf and dictionary learning. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2296–2303 (2012)
- Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: *Advances in Neural Information Processing Systems*, pp. 545–552 (2006)
- Hou, X., Zhang L.: Saliency detection: a spectral residual approach. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (2007)
- Cheng, M.M., Zhang, G.X., Mitra, N.J., Huang, X., Hu, S.M.: Global contrast based salient region detection. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 409–416 (2011)
- Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1597–1604 (2009)
- Achanta, R., Susstrunk, S.: Saliency detection using maximum symmetric surround. *17th IEEE International Conference on Image Processing (ICIP)*, pp. 2653–2656 (2010)
- Goferman, S., Zelnik-Manor, L., Tal, A.: Context-aware saliency detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(10), 1915–1926 (2012)
- Zhang, H., Xu, M., Zhuo, L., Havyarimana, V.: A novel optimization framework for salient object detection. *Vis. Comput.* (2014). doi:10.1007/s00371-014-1053-z
- Zhai, Y., Shah, M.: Visual attention detection in video sequences using spatiotemporal cues. In: *Proceedings of the 14th Annual ACM International Conference on Multimedia*, pp. 815–824 (2006)
- Chang, K.Y., Liu, T.L., Chen, H.T., Lai, S.H.: Fusing generic objectness and visual saliency for salient object detection. In: *IEEE International Conference on Computer Vision (ICCV)*, pp. 914–921 (2011)
- Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M. H.: Saliency detection via graph-based manifold ranking. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3166–3173 (2013)
- Jiang, B., Zhang, L., Lu, H., Yang, C., Yang, M. H.: Saliency detection via absorbing Markov chain. In: *IEEE International Conference on Computer Vision*, pp. 1665–1673 (2013)
- Toet, A.: Computational versus psychophysical bottom-up image saliency: a comparative evaluation study. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(11), 2131–2146 (2011)

21. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels. Technical report, EPFL, Tech. Rep. 149300 (2010)
22. Aldous, D., Fill, J.: Reversible Markov chains and random walks on graphs. <http://www.stat.berkeley.edu/~aldous/RWG/book.pdf>
23. Grinstead, C.M., Snell, J.L.: Introduction to probability, pp 10–125. American Mathematical Society (1998)
24. Norris, J.R.: Markov Chains. Cambridge University Press, Cambridge (1998)
25. Otsu, N.: A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **9**, 62–66 (1979)
26. Einhäuser, W., König, P.: Does luminance contrast contribute to a saliency map for overt visual attention? *Eur. J. Neurosci.* **17**(5), 1089–1097 (2003)
27. Judd, T., Ehinger, K., Durand, F., Torralba, A.: Learning to predict where humans look. In: *IEEE 12th International Conference on Computer Vision*, pp. 2106–2113 (2009)
28. Wei, Y., Wen, F., Zhu, W., Sun, J.: Geodesic saliency using background priors. In: *Computer Vision-ECCV*, pp. 29–42. Springer, Berlin (2012)
29. Martinez-Gil, J., Aldana-Montes, J.F.: Semantic similarity measurement using historical google search patterns. *Inf. Syst. Front.* **15**(3), 399–410 (2013)
30. Tan, P.N., Steinbach, M., Kumar, V.: *Introduction to Data Mining*, pp. 500–502. Pearson Addison-Wesley, Boston (2005)
31. Jiang, H., Wang, J., Yuan, Z., Wu, Y., Zheng, N., Li, S.: Salient object detection: a discriminative regional feature integration approach. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2083–2090 (2013)
32. Batra, D., Kowdle, A., Parikh, D., Luo, J., Chen, T.: Interactively co-segmenting topically related images with intelligent scribble guidance. *Int. J. Comput. Vis. (IJCV)* **93**(3), 273–292 (2011)
33. Batra, D., Kowdle, A., Parikh, D., Luo, J., Chen, T.: iCoseg: interactive co-segmentation with intelligent scribble guidance. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3169–3176 (2010)
34. Alpert, S., Galun, M., Basri, R., Brandt, A.: Image segmentation by probabilistic bottom-up aggregation and cue integration. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (2007)



Wenjie Zhang is currently pursuing Ph.d. degree in College of Automation, Chongqing University, Chongqing, China.



Qingyu Xiong is the dean of School Software Engineering, Chongqing University, Chongqing 400044, China. His current research interests include the areas of computer vision and pattern recognition with focus on visual tracking, and saliency detection.



Weiren Shi is the professor of College of Automation, Chongqing University, Chongqing, China.



Shuhan Chen PH.D., works at College of Information Engineering, Yangzhou University, Yangzhou, China.