Pattern Recognition 64 (2017) 39–50

Contents lists available at ScienceDirect

Pattern Recognition

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

A novel graph-based optimization framework for salient object detection

Jinxia Zhang[a,b,*], Krista A. Ehinger[c,d], Haikun Wei[a,*], Kanjian Zhang[a], Jingyu Yang[b]

[a] Key Laboratory of Measurement and Control of CSE, Ministry of Education, School of Automation, Southeast University, Nanjing 210096, China
[b] Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, Nanjing University of Science and Technology, Nanjing 210094, China
[c] Brigham and Women's Hospital, Cambridge, MA02139, United States
[d] Harvard Medical School, Cambridge, MA02139, United States

ARTICLE INFO

Keywords:
Optimization framework
Multiple graphs
Visual rarity
Saliency detection

ABSTRACT

In traditional graph-based optimization framework for salient object detection, an image is over-segmented into superpixels and mapped to one single graph. The saliency value of each superpixel is then computed based on the similarity between connected nodes and the saliency related queries. When applying the traditional graph-based optimization framework to the salient object detection problem in natural scene images, we observe at least two limitations: only one graph is employed to describe the information contained in an image and no cognitive property about visual saliency is explicitly modeled in the optimization framework. In this work, we propose a novel graph-based optimization framework for salient object detection. Firstly, we employ multiple graphs in our optimization framework. A natural scene image is usually complex, employing multiple graphs from different image properties can better describe the complex information contained in the image. Secondly, we model one popular cognitive property about visual saliency (visual rarity) in our graph-based optimization framework, making this framework more suitable for saliency detection problem. Specifically, we add a regularization term to constrain the saliency value of each superpixel according to visual rarity in our optimization framework. Our experimental results on four benchmark databases with comparisons to fifteen representative methods demonstrate that our graph-based optimization framework is effective and computationally efficient.

1. Introduction

Since Itti et al. [1] introduced the first computational model of visual saliency, a large number of methods have been proposed for salient object detection in images [2–18]. For detailed reviews, see [19,20]. Salient object detection can be utilized as a preprocessing operation in many applications of computer vision, including image quality assessment [21,22], image segmentation [23], image and video compression [24,25], picture collage [26,27], video summarization [28,29], object detection and recognition [30–32], visual tracking [33,34], and content based image retrieval [35]. Therefore, the performance of salient object detection algorithms is very important for subsequent applications.

Recently, a few researchers have proposed salient object detection methods using a graph-based optimization framework. Since the graph is able to conveniently describe the structure information within an image, these saliency detection methods achieve top performances. Yang et al. [13] employ a graph ranking based optimization framework for salient object detection in images. This work focuses on the computation of the saliency-related cues or queries and a two-stage scheme is proposed which considers both the foreground and the background queries. Lu et al. [14] propose a method to learn optimal queries for salient object detection. In this work, both the bottom-up saliency and a set of mid-level vision features are computed for each image region. Then the combination of features is learned using a large-margin formulation to represent the optimal queries. In the work [16], Li et al. propose a regularized random walk ranking framework to formulate pixel-wise saliency based on the superpixel-based foreground and background saliency estimations. Gong et al. [17] exploit an iterated optimization algorithm to detect saliency from simple regions to difficult regions using the teaching-to-learn and learning-to-teach strategies.

In the optimization frameworks employed by the above methods, we observe that these frameworks are usually based on two kinds of information: similarity between the connected nodes (pixels or superpixels) in the constructed graph and the saliency-related queries. In this paper, we propose a novel graph-based optimization framework to incorporate not only the information already used by existing frame-

* Corresponding author at: School of Automation, Southeast University, 2 Sipailou Street, Nanjing, Jiangsu 210094, China.
E-mail address: hkwei@seu.edu.cn (H. Wei).

http://dx.doi.org/10.1016/j.patcog.2016.10.025
Received 26 July 2016; Received in revised form 14 September 2016; Accepted 19 October 2016
Available online 22 October 2016

works but also other important priors for salient object detection, e.g. cognitive properties of visual saliency and cognitive properties of perceptual grouping.

Cognitive properties of visual saliency were incorporated into the first computational model of saliency detection [1], and many early algorithms inspired by this work consider similar cognitive properties. The most frequently used property is visual rarity. The idea behind visual rarity is that human attention is usually drawn to image regions with rare or unusual features, and not drawn to regions with features common in that image. Mancas et al. [3] model visual rarity based on information theory and use this to detect visual saliency. Hou and Zhang [5] extract the rare parts of an image in the frequency domain and use inverse Fourier transform to select these regions in the spatial domain. Another useful cognitive property of visual saliency adopted by salient object detection algorithms is a central bias prior. This prior assumes that the central part of an image is more likely to be a salient foreground object while the borders of an image are more likely to be background. Zhu et al. [36] model this prior by assuming that an image background region is usually heavily connected to the image boundary. Jiang et al. [37] choose the boundary nodes as the absorbing nodes in a Markov chain and compute a saliency value for each node based on the absorption time from this node to the boundary absorbing nodes. Besides cognitive properties of visual saliency, cognitive properties of perceptual grouping are also employed by the existing saliency detection algorithms. One of the strongest perceptual grouping properties is spatial proximity, which has been employed extensively in computer vision research [13,37,38]. This property states that items which are near each other are more likely to group together as one object than items that are more widely separated. Yang et al. [13], Jiang et al. [37] and Gopalakrishnan et al. [39] make use of spatial proximity to construct a graph where each node is connected to its spatial neighbors in the image. This reflects the fact that pixels which are spatially close are likely to have similar saliency values. Another commonly used perceptual grouping property is color similarity [40,41]. This property states that image regions with similar colors are more likely to group together, probably reflecting the fact that real-world objects of the same kind usually have similar appearance. Achanta et al. [40] measure the saliency value of each pixel by computing the magnitude of the difference between the color features of this pixel and the mean color features of the image.

By studying the existing optimization frameworks, we observe that no cognitive properties of visual saliency are explicitly modeled in these classical graph-based optimization frameworks. Based on this observation, we model one popular visual saliency property (visual rarity) explicitly in our graph-based optimization framework. Specifically, we model visual rarity as a regularization term in the graph-based optimization framework. In this way, we can constrain the saliency value of each image region according to the visual rarity property. Another observation about the existing optimization frameworks is that usually only one graph is constructed for salient object detection. However, constructing only one graph may neglect important structure information contained in complex natural scene images. In this paper, we propose to employ multiple graphs to describe the image information from different image properties. Inspired by previous graph-based methods [13,14,17], we construct one graph based on spatial location features. Since the above mentioned spatial proximity property and central bias prior provide location related information about saliency detection, we construct the first graph according to these two cognitive properties: two nodes are connected if they are spatially adjacent to indicate that nodes with close locations are likely to belong to the same class and if both of them are boundary superpixels to indicate that nodes on the image border tend to belong to the background. Besides the graph based on location features, we construct another graph based on color features inspired by color similarity property to reflect the fact that image regions with similar appearance are likely to be part of the same class (foreground or background). This property can help to detect multiple salient foreground objects or a single object with a complex pattern.

The main contributions of our paper are summarized as follows: We propose a novel graph-based optimization framework for salient object detection. Firstly, we employ multiple graphs and use a weighted combination in the optimization framework. We construct separate graphs based on spatial features and color features to better capture the complex information in a natural scene image. Secondly, we explicitly model visual rarity as a regularization term in our graph-based optimization framework to better detect visual salient objects.

We test our graph-based optimization framework on four different saliency databases and demonstrate that our method outperforms fifteen state-of-the-art methods, including previous graph-based optimization methods. The remainder of this paper is organized as follows: the proposed graph-based optimization framework is introduced in Section 2. Section 3 describes how we construct multiple graphs and acquire the regularization term based on visual rarity. Section 4 provides our method to compute the queries of the salient objects. Section 5 describes our testing procedure and presents the comparisons to other salient object detection methods on different saliency databases, and Section 6 concludes the paper.

2. The proposed framework

In a graph-based optimization framework, the goal is set to find a function F: $V \Rightarrow \mathbb{R}$ which assigns a value $F(i)$ to each node $V(i)$. Given $V = \{V(1), \ldots, V(n)\}$, some of the nodes are queries. Q is defined such that if $V(i)$ is a query $Q(i) = 1$, and otherwise $Q(i) = 0$. Let W: $V*V \Rightarrow \mathbb{R}$ denote the weighted adjacency matrix for a graph in which the weight of each edge is defined as the similarity of the connected nodes. Different function values F of different nodes are to be computed based on their similarities to the connected nodes and the queries [42]. In traditional graph-based optimization framework, the optimal function (F^*) based on the graph Laplacian can be computed by solving the following optimization problem:

$$F^* = \arg\min_F \{\mathcal{S}(F) + \mu C(F)\}\mathcal{S}(F) = \frac{1}{2}\sum_{i,j=1}^{n}(W(ij)\|F(i) - F(j)\|^2)C(F)$$
$$= \sum_{i=1}^{n}(\|F(i) - Q(i)\|^2) \quad (1)$$

In the term $\mathcal{S}(F)$, $F(i)$ and $F(j)$ are values for the nodes $V(i)$ and $V(j)$ separately. From the formula, we can see that this term indicates the smoothness constraint for the graph. If two nodes $V(i)$ and $V(j)$ are connected with a high weight, the values of these two nodes should be close. In the fitting constraint term $C(f)$, $F(i)$ and $Q(i)$ indicate the function value and the query value of the node $V(i)$. So this term represents the amount of weight put on the initial query assignment. $\mu \in (0, \infty)$ is a parameter, which specifies the relative balance of these two terms.

Recently, this optimization framework has been applied to the problem of salient object detection. Some researchers focus on query acquisition [13,14]. Some researchers focus on the graph construction [15]. And some researchers place emphasis on the computation sequence given the labeled nodes (queries) and unlabeled nodes [17]. These methods are among the state-of-the-art algorithms for salient object detection because graph-based optimization framework facilitates the description of the structure information in an image and the fusion of the priors via the queries.

By observing the classical graph-based optimization framework, we find that it has two limitations when applied to salient object detection in natural scene images. Firstly, only one graph is employed in this framework. Natural scene images tend to have complex relations between image regions because there are diverse objects with any numerosity and any position in the image. Using only one graph to represent the structure of an image may omit important information

for saliency detection. Secondly, this framework does not explicitly represent the cognitive properties of visual saliency in the graph-based optimization function. As is introduced in Formula (1), the classical framework makes use of the information between connected nodes in the smoothness constraint term $S(F)$ and the information of the queries in the term $C(f)$. In this framework, no cognitive property of visual saliency is included. So the performance of this framework is limited when it is applied to salient object detection. In this paper, we put our emphasis on the graph-based optimization problem per se and introduce a novel graph-based optimization framework to overcome the above-mentioned limitations. Our proposed graph-based optimization framework is formulated as follows:

$$\begin{aligned} F^* &= \arg\min_F \left\{ \frac{1}{2}\sum_{k=1}^{m} \alpha_k \mathcal{S}_k(F) + \mu C(F) + \gamma \mathcal{R}(F) \right\} \mathcal{S}_k(F) \\ &= \sum_{i,j=1}^{n} (W_k(ij)\|F(i) - F(j)\|^2) C(F) = \sum_{i=1}^{n} (\|F(i) - Q(i)\|^2) \mathcal{R}(F) \\ &= \sum_{i=1}^{n} (O(i)\|F(i)\|^2) \end{aligned} \tag{2}$$

Firstly, we exploit multiple graphs to better describe the structure information between different image regions and include multiple smoothness constraint terms $\mathcal{S}_k(F)$ in the optimization framework based on these graphs. We employ m weighted adjacency matrixes to represent m graphs which describe different relationships between image regions based on different principles. We then use the parameter α_k to compute a weighted combination of different smoothness constraint terms $\mathcal{S}_k(F)$. The sum of these parameters $\sum_{k=1}^{m} \alpha_k$ is equal to 1. By combining these terms together, we can measure the structure of an image from different image properties.

Our second contribution is that we explicitly model one of the well-known cognitive properties of visual saliency (i.e., visual rarity) in our graph-based optimization framework to better detect visually salient objects in images. We introduce a regularization term $\mathcal{R}(f)$ in addition to the commonly used smoothness constraint term $S(f)$ and fitting constraint term $C(f)$ in the optimization framework. We call this regularization term the rarity term in our paper since it is constructed based on visual rarity. Visual rarity captures the fact that human eyes are often attracted to the rare features in an image but not to the common features. Some researchers have modeled visual rarity for saliency detection. For example, Mancas et al. [3] propose a visual saliency detection model based on rarity and measure saliency using information theory. In this paper, we model the visual rarity property in our graph-based optimization framework as a regularization term. In this regularization term, we introduce a vector $O(i)$ indicating the occurrence rate or frequency with which the feature value of a node $V(i)$ is present in the image. If a node has common features, this node should have a large occurrence rate and thus should have a low saliency value. We use a parameter γ to represent the weight of this regularization term.

We can see that there are three important factors for our graph-based optimization framework: multiple graphs W_k for the smoothness constraint terms $\mathcal{S}_k(F)$, the queries Q for the fitting constraint term $C(f)$, and the occurrence rate O for the rarity term $\mathcal{R}(f)$. All of these factors are critical in determining the final saliency values. In Section 3, we describe how we calculate multiple graphs W_k and the occurrence rate O using important priors for salient object detection. And in Section 4, we introduce how we compute salient queries Q based on a central bias prior and the consistency between connected image regions.

3. Multi graph construction and rarity term acquisition

In our method, we first use the SLIC approach [43] to over-segment an image into homogeneous superpixels and then define these homogeneous superpixels as the nodes of each graph. The number of superpixels in each image is set to 200 in our method, as in previous graph-based works [13,37]. The spatial and color features of each superpixel are defined as its mean in the image coordinates and CIELab color space, respectively. The values of the spatial and color features are normalized to be in [0, 1].

3.1. Multi-graph construction

As described in Section 2, connected nodes with large edge weights would have similar saliency values. So, well-constructed graphs should connect the nodes which are likely to be in the same class (foreground or background) and weight these edges appropriately. Although well-constructed graphs are very important to determining the final saliency values, most previous graph-based salient object detection methods do not pay much attention to graph construction. These methods construct a single graph to describe the relationships between different image regions. Since a natural scene image tends to have complex structure between image regions, using only one graph may omit important information about the image. In this paper, we propose the construction of multiple graphs to describe image information from different image feature spaces. Specifically, we construct two graphs based on spatial locations and color features, respectively. The following sections present how we choose edges and compute edge weights for these graphs.

3.1.1. Choosing edges for different graphs

In previous graph-based salient object detection methods [13,37], the graph is constructed by choosing the edges based on spatial location features. Inspired by these works, we employ a similar method to choose edges for our first graph. Since spatial proximity property and central bias prior provide spatial location related information about saliency detection, we choose edges for our first graph based on these properties. According to spatial proximity property, the image regions which are spatially nearby tend to belong to the same class. Thus, we connect each node in this graph to its spatial neighbors (the superpixels which share a boundary) and to its spatial neighbors' neighbors (the superpixels which share a boundary with any node in the first set of spatial neighbors). Inspired by central bias prior, regions on the image borders are likely to be background [44,45]. We further fully connect the nodes on the borders of the image to reflect the prior that they tend to belong to the "background" class. We introduce a weighted adjacency matrix W_1 to represent our first constructed graph based on spatial location:

$$W_1(i,j) = \begin{cases} A(i,j) & \text{if } j \in N_s(i) \\ A(i,j) & \text{if } i \in \mathbb{B} \& j \in \mathbb{B} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

We use the similarity between the connected nodes i and j (denoted as $A(i,j)$) to represent the weight of the corresponding edge. $N_s(i)$ denotes the spatial neighbors and the neighbors' neighbors of node i. $\mathbb{B}$ represents the node set which includes the nodes on the image borders.

Unlike previous graph-based methods [13,37] which employ only a single graph, we exploit multiple graphs for salient object detection. We construct another graph based on the cognitive property of color similarity. According to this property, image regions with similar colors are likely to share the same label. Specifically, we connect the image regions which are neighbors in the color space to construct this graph. This kind of graph helps to detect multiple salient foreground objects or a single object with a complex pattern in an image. The weighted adjacency matrix W_2 for this graph is formulated as follows:

$$W_2(i,j) = \begin{cases} A(i,j) & \text{if } D_c(i,j) < \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

In this formula, $D_c(i,j)$ denotes the color distance between the

Fig. 1. An illustration of the occurrence rate.

nodes i and j, which is defined as the Euclidean distance between the color features. We set ε as 0.15 in our method.

3.1.2. Computing edge weights

In the previous section, we have introduced how we choose different edges for different graphs. In this section, we introduce how to compute the weight between connected nodes i and j based on the similarity $A(i, j)$ of different nodes for these graphs.

In existing graph-based salient object detection methods, edges are usually weighted by the color distance between the nodes [13,37]:

$$A(i,j) = \exp\left(-\frac{D_c(i,j)}{\sigma^2}\right) \tag{5}$$

σ is a scale parameter which controls the strength of the distance. If σ has a small value, only nodes with close features would make a contribution. If σ has a large value, the nodes with larger distances would also have influence on each other. We experimentally test the effect of this parameter in Section 5.

This definition of the edge weight may produce false alarms if some part of the distant background has a very similar color to the foreground. Based on this consideration, we compute the edge weight using not only the color distance but also the spatial distance [15]:

$$A(i,j) = \exp\left(-\frac{D_c(i,j) + D_s(i,j)}{2\sigma^2}\right) \tag{6}$$

$D_s(i, j)$ represents the spatial distance between the nodes i and j. A popular way to compute the spatial distance is to compute the Euclidean distance in the image coordinates. However, there is a drawback when applying the Euclidean spatial distance to the problem of salient object detection: the Euclidean spatial distance is largest between the opposite borders of an image (e.g., the left and right borders of an image), which indicates that there is a strong tendency for the opposite image borders to be assigned to different classes. In fact, different image borders are likely to belong to the same class ("background"). To solve this problem, we employ the sine spatial distance proposed in our previous work [15] to replace the Euclidean spatial distance:

$$D_s(i,j) = \sqrt{(\sin(\pi \cdot |x_i - x_j|))^2 + (\sin(\pi \cdot |y_i - y_j|))^2} \tag{7}$$

In this formula, x_i and y_i represent the horizontal and vertical axes of a node i in the image plane, which have been normalized to be in [0, 1]. $\sin(\pi \cdot |x_i - x_j|)$ computes the sine spatial distance between two nodes i and j along the horizontal coordinate. Similarly, $\sin(\pi \cdot |y_i - y_j|)$ is the sine spatial distance along the vertical coordinate. Based on this definition, the nodes with a small Euclidean spatial distance would have a small sine spatial distance. Also, the nodes at the opposite borders of the image would be considered spatially close, which is more likely to make the image borders share the same label.

3.2. Rarity term acquisition

In our approach, we model visual rarity as a regularization term in the graph-based optimization framework. We introduce a vector O in this term as is introduced in Formula (2). This vector records the occurrence rate or frequent rate $O(i)$ that the features of a node i occur in the image.

Since saliency detection is usually utilized as a preprocessing operation for other computer vision applications, saliency detection methods should be efficient. To this end, we simply compute the occurrence rate O based on our already constructed graphs, which is defined in the following formula:

$$O(i) = \sum_{j=1}^{n} W_1(i,j) + \sum_{j=1}^{n} W_2(i,j) \tag{8}$$

In the first half part of this formula, we include the degree of the node i in the graph W_1. According to the visual rarity, image background tends to occupy more area than salient foreground in an image and should have a higher occurrence rate O. Our graph W_1 is constructed so that the nodes on the image borders are connected to more nodes and therefore have larger degree ($\sum_{j=1}^{n} W_1(i,j)$). So, including this part in Eq. (8) makes the background nodes on the image borders have high occurrence rates.

We also include the degree of the node i in the second graph W_2. Inspired by the visual rarity, salient foreground tends to have rare color features and connect to fewer nodes in graph W_2. So, adding $\sum_{j=1}^{n} W_2(i,j)$ in Eq. (8) would make the salient foreground nodes have lower occurrence rates. As an illustration, we diagram the occurrence rate O in Fig. 1. It can be seen that the background in an image which usually has frequent features will have high occurrence rate while the salient foreground which usually has rare features will have low

occurrence rate. From this figure, we also note that this rarity term can help to detect the salient object near the image borders (second image) and multiple salient objects (third image).

4. Query computation

In this section, we introduce how we compute the salient queries based on our constructed graphs and the central bias prior of visual saliency. According to the central bias prior, the nodes on the image borders are more likely to be background. We use each of the four image borders separately to get four temporal query results based on the following formula:

$$Q^* = \arg\min_{Q}\left\{\frac{1}{2}\sum_{k=1}^{2}\alpha_k S_k(Q) + \mu C(Q)\right\}S_k(Q) = \sum_{i,j=1}^{n}(W_k(ij)\|Q(i)-Q(j)\|^2)C(Q) = \sum_{i=1}^{n}(\|Q(i)-B(i)\|^2) \quad (9)$$

In this formula, $B(i)$ indicates whether a node belongs to one of the four image borders. If a node i belongs to the image border, $B(i) = -1$ indicates that this node is a background node, and otherwise $B(i) = 0$. Each of the four temporal query results is computed based on the relationship between different nodes as defined by our constructed graphs and one of the image borders. By setting the derivative of the above function to be zero, the temporal query result is:

$$Q^* = [\alpha_1(D_1 - W_1) + \alpha_2(D_2 - W_2) + \mu I]^{-1}B \quad (10)$$

W_1 and W_2 are our two constructed graphs. D_1 is defined as $\sum_{j=1}^{n} W_1(i,j)$, representing the degree of each node for W_1. Similarly, D_2 is defined as $\sum_{j=1}^{n} W_2(i,j)$, representing the degree of each node for W_2. According to Eq. (10), the nodes which have similar features to the image borders would have a value close to −1, and otherwise would have a value close to 0. Based on four image borders (B_{top}, B_{down}, B_{left} and B_{right}) separately, we can get four temporal query results (Q_{top}, Q_{down}, Q_{left} and Q_{right}). We then normalize and multiply them to get the final queries:

$$Q_{final} = Norm(Q_{top}).*Norm(Q_{down}).*Norm(Q_{left}).*Norm(Q_{right}) \quad (11)$$

In this formula, $Norm(\cdot)$ means to normalize the result to be in [0, 1]. We use node multiplication to combine different temporal query results.

After getting the final queries in this section, the constructed graphs and the rarity term in the previous section, we can compute the saliency values of different nodes based on Eq. (2). By setting the derivative of this formula to be 0, we get the final saliency result:

$$F^* = [\alpha_1(D_1 - W_1) + \alpha_2(D_2 - W_2) + \gamma(D_1 + D_2) + \mu I]^{-1}Q_{final} \quad (12)$$

The normalized result of F^* indicates the likelihood that a node belongs to the salient foreground. The procedure of our proposed salient object detection method is summarized in Algorithm 1.

Algorithm 1. Salient object detection based on a novel graph-based optimization framework.

Input: An image and required parameters.

1. Construct two graphs based on spatial locations and color features, respectively: Choose graph edges by Eq. (3) for G_1 and choose graph edges by Eq. (4) for G_2; Compute edge weights by Eqs. (6) and (7) for these two graphs.
2. Compute the occurrence rate O by Eq. (8).
3. Compute four temporal query results Q_{top}, Q_{down}, Q_{left} and Q_{right} by Eq. (10). Then get the final query result Q_{final} by Eq. (11).
4. Compute the final saliency result by Eq. (12) according to the constructed graphs, the occurrence rate, and the final query result.

Output: A saliency map.

5. Experimental comparisons

In this section, we compare our proposed method to fifteen state-of-the-art saliency detection methods: IT [1], LC [46], FT [40], HC [47], GB [48], RC [47], CB [49], HM [50], HS [51], AM [37], GR [13], BD [36], CL [17], RR [16] and GP [52]. Following [40,47], we chose different methods based on various principles: recency (CL, RR and GP) and variety (IT is biologically motivated; LC is purely computational; FT is frequency tuned; HC is based on color histogram; RC uses regional contrast; CB is based on context and shape knowledge; and GB, AM, GR, CL, RR and GP are graph based).

We evaluate these methods on four salient object databases: the 10 000-image MSRA10K database [47], the 643-image iCoseg database [53], the 1000-image Extended Complex Scene Saliency Database (ECSSD) [51] and the 300-image SOD database [54]. In the MSRA10K database, each image has an unambiguous salient object. The iCoseg database contains images with one or multiple foreground objects. The ECSSD database contains natural images with complex foreground and background patterns. The SOD database contains 300 challenging images from the Berkeley segmentation database [55].

5.1. Evaluation metrics

We evaluate different methods using standard precision–recall curves [56]. To get the precision–recall curve for each method, the saliency map is binarized at each threshold in the range [0:1:255] and the precision and recall values at each threshold are computed by comparing the binary map with the ground truth. Let F_{gt} and F_d denote the foreground of the ground truth and the detected foreground in the binary map, respectively. The precision and recall values are mathematically defined as:

$$Precision = area(F_d \cap F_{gt})/area(F_d) Recall = area(F_d \cap F_{gt})/area(F_{gt}) \quad (13)$$

In many cases, high precision and recall values are both required. So, as measures of overall performance we report the area under the precision–recall curve (average precision, AP) [56] and the F-measure [57] using twice the mean value of the saliency map as a threshold. The formulation of F-measure is as follows and we set $\beta^2 = 1$ to place equal emphasize on precision and recall:

$$F_\beta = \frac{(1+\beta^2)\cdot precision \cdot recall}{\beta^2 \cdot precision + recall} \quad (14)$$

5.2. Experimental setup

There are four primary parameters in our proposed method: three parameters (α_1, μ and γ) in Eqs. (10) and (12) and one parameter (σ) in Eq. (6). The parameter α_2 in Eqs. (10) and (12) is set according to $\sum_{k=1}^{2}\alpha_k = 1$. We test different values for these parameters on the 300-image SOD database using average precision (AP).

The parameters α_1 and α_2 specify the relative weights of two constructed graphs, respectively, where $\alpha_1 \in [0, 1]$ and $\alpha_2 = 1 - \alpha_1$. If $\alpha_1 = 0$, the result only depends on the second graph. If $\alpha_1 = 1$, the result only depends on the first graph. From Table 1, we can see that the performance is better when combining two graphs than only using one graph and the performance is best when the weight $\alpha_1 = 0.4$. So in this paper, we set $\alpha_1 = 0.4$ and $\alpha_2 = 0.6$ for all experiments on different databases.

The parameter $\mu \in (0, \infty)$ provides the contribution of the fitting term $C(F)$. The larger μ is, the more $C(F)$ will contribute to the result. In this paper, we set $\mu = 0.01$ for all experiments on different databases

Table 1
Performance comparison with different α_1.

α_1	0	0.1	0.2	0.3	**0.4**	0.5
AP	0.6395	0.7063	0.7244	0.7320	**0.7368**	0.7315
α_1	0.6	0.7	0.8	0.9	1	–
AP	0.7266	0.7197	0.7127	0.7031	0.7008	–

The optimal performance is highlighted with bold fonts.

Table 2
Performance comparison with different μ.

μ	0.001	**0.01**	0.02	0.03	0.04	0.05
AP	0.7300	**0.7368**	0.7291	0.7264	0.7217	0.7210

The optimal performance is highlighted with bold fonts.

Table 3
Performance comparison with different γ.

γ	0.001	0.01	**0.02**	0.03	0.04	0.05
AP	0.7181	0.7322	**0.7368**	0.7310	0.7315	0.7308

The optimal performance is highlighted with bold fonts.

based on Table 2.

Similarly, the parameter $\gamma \in (0, \infty)$ provides the contribution of the rarity term $\mathcal{R}(f)$. Larger γ means that $\mathcal{R}(f)$ will contribute more to the result. We set $\gamma = 0.02$ for all experiments in this paper based on Table 3.

Table 4 shows the effect when changing the value of σ^2. The results are satisfactory when σ^2 is set from 0.04 to 0.1. In our implementation, we set $\sigma^2 = 0.05$ for all experiments on different databases.

5.3. Quantitative comparison

In this section, we compare our method with fifteen state-of-the-art methods on four different databases with multiple metrics.

MSRA10K: Fig. 2(a) shows the precision–recall curves of different methods. It can be seen that the proposed method outperforms GR [13], CL [17], RR [16] and GP [52] which are also graph-based methods and have top performances. We also compute the average precision and *F*-measure for different methods. Fig. 2(b) and (c) shows that our proposed method achieves the highest average precision and *F*-measure values. Overall, the quantitative results using three metrics demonstrate that the proposed method outperforms the state-of-the art methods on this large database.

iCoseg: We further evaluate the proposed method on the iCoseg database in which each image contains one or multiple foreground objects, shown in Fig. 2(d)–(f). We note that the performances of different methods drop a little in comparison to the performances on the MSRA10K database. This indicates that it is harder to detect all of the foreground regions when there are multiple salient objects in an image. From the results, it can be seen that the proposed method outperforms other methods on this database. It demonstrates that the use of multiple graphs representing different image properties and the rarity term in our proposed method can help to detect all salient objects in the image.

Table 4
Performance comparison with different σ^2.

σ^2	0.01	0.02	0.03	0.04	**0.05**	0.06
AP	0.3167	0.3276	0.6339	0.7283	**0.7368**	0.7326
σ^2	0.07	0.08	0.09	0.10	0.15	0.20
AP	0.7297	0.7268	0.7228	0.7231	0.7165	0.7135

The optimal performance is highlighted with bold fonts.

ECSSD: The ECSSD database is the extended complex scene saliency database, which contains 1000 natural images with complex foreground and background patterns. The performance comparison in Fig. 2(g)–(i) demonstrates that the proposed method can better detect salient objects in complex natural images than other state-of-the-art methods.

SOD: The SOD database is a difficult salient object detection database, which contains a number of natural images with multiple foreground objects and complex patterns. The performances of different methods are shown in Fig. 2(j)–(l). Fig. 2(j) shows that when the precision value is equal, our proposed method has a larger recall value than other methods. Fig. 2(k) and (l) demonstrates that the proposed method outperforms other methods on the AP and *F*-score measures. We also note that the performances of different methods on this database are much lower than the performances on the MSRA10K database. This indicates that there is room for improvement in future methods to better detect salient objects in natural images with multiple foreground objects and complex patterns.

5.4. Visual comparison

We present some saliency maps generated by our method as well as fifteen other methods for visual comparison in Fig. 3. Representative natural images have been chosen to highlight the differences between different methods. The first and second images contain a complex foreground object. Most methods highlight part of the salient object (e.g. the roof part of the building) while our proposed method can uniformly highlight the whole salient object. The third image contains complex background. Many methods would wrongly highlight part of the background, but our proposed method is more accurate because we introduce the rarity term in our graph-based optimization framework, which can help to inhibit the image background. The forth image contains one small salient object. A lot of methods wrongly highlight the smoke near the plane. The reason why our method can give a better saliency detection result is that the rarity term in our framework helps to highlight the small red plane and inhibit the background. The fifth image contains multiple small salient objects. Our method is able to detect all of them. The last image also contains multiple salient objects. Among these salient objects, part of them are not close to others. Some methods (e.g. AM, GR and RR, which are also graph-based methods) cannot uniformly highlight all the objects. The major reason why our method works better here is that we include the color-based graph in our method, which can help to highlight all the salient objects even if they are not close to each other. In a word, the saliency maps generated by our proposed method can better highlight all the salient objects with fewer noisy results.

5.5. Examination of design options

We examine different design options on the SOD database, shown in Fig. 4. We first examine the major innovations of the proposed method in Fig. 4(a). The yellow curve provides the performance of the method with one graph connecting nodes based on spatial locations only and without the rarity term. Similarly, the black curve provides the performance of the method with one graph based on color features only and without the rarity term. The blue curve gives the performance of the method with two graphs (one graph is constructed by connecting

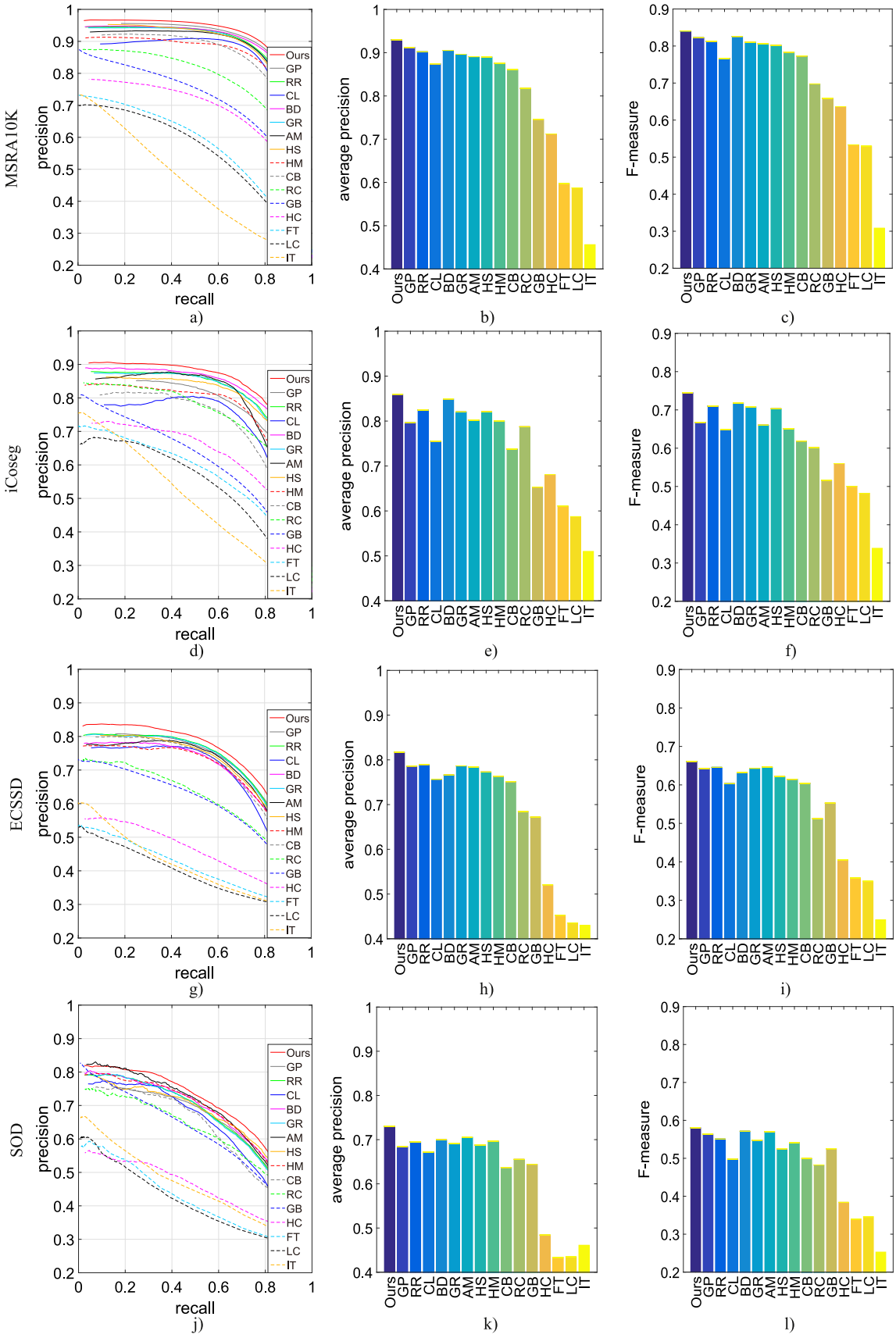

Fig. 2. The PR curve, average precision and *F*-measure for different methods on MSRA10K, iCoseg, ECSSD and SOD databases.

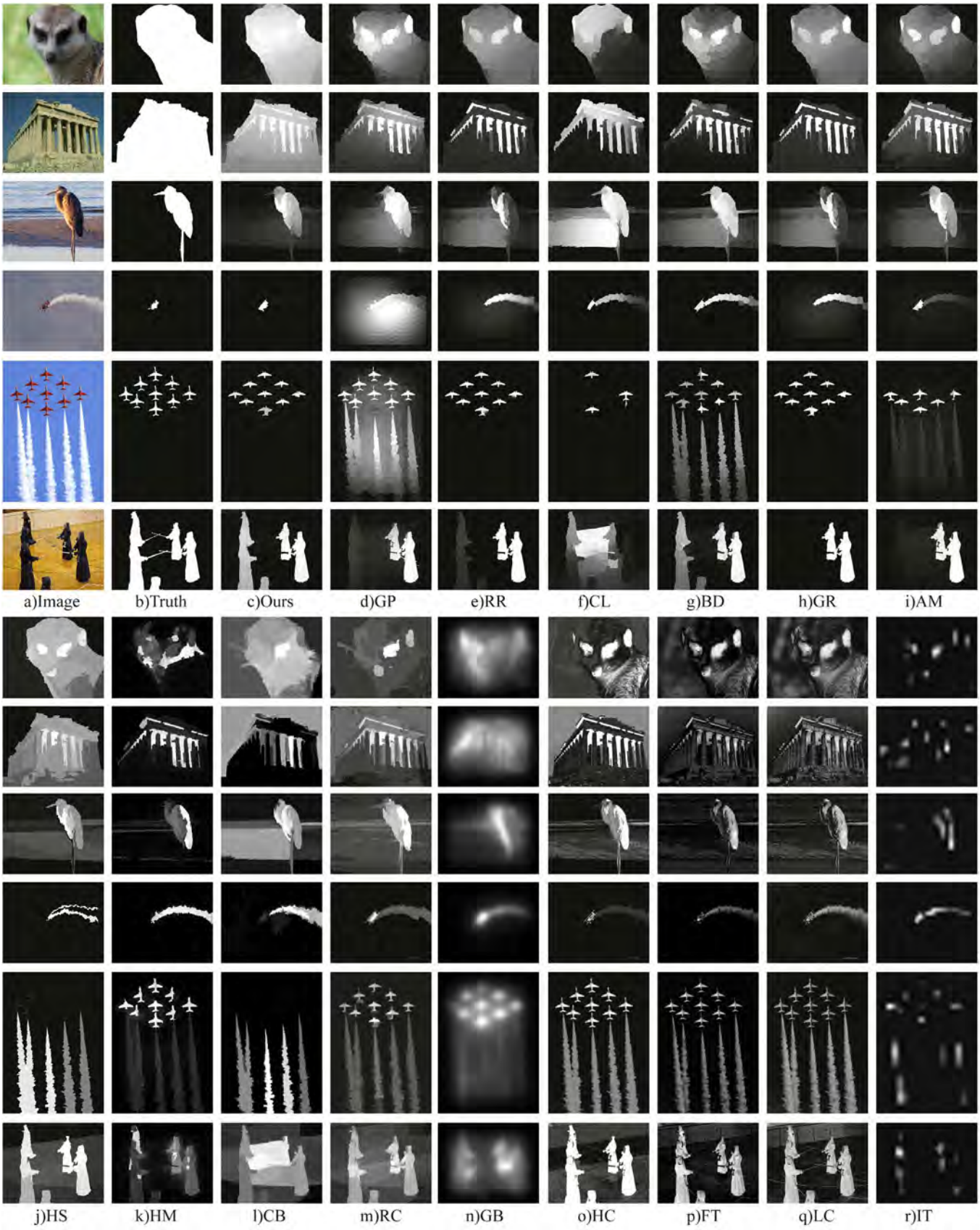

Fig. 3. Comparison of different salient object detection methods. The column (a) is the original image, the column (b) is the ground truth, and the remaining columns are results of the evaluated methods. Our method is the column (c) of the results. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

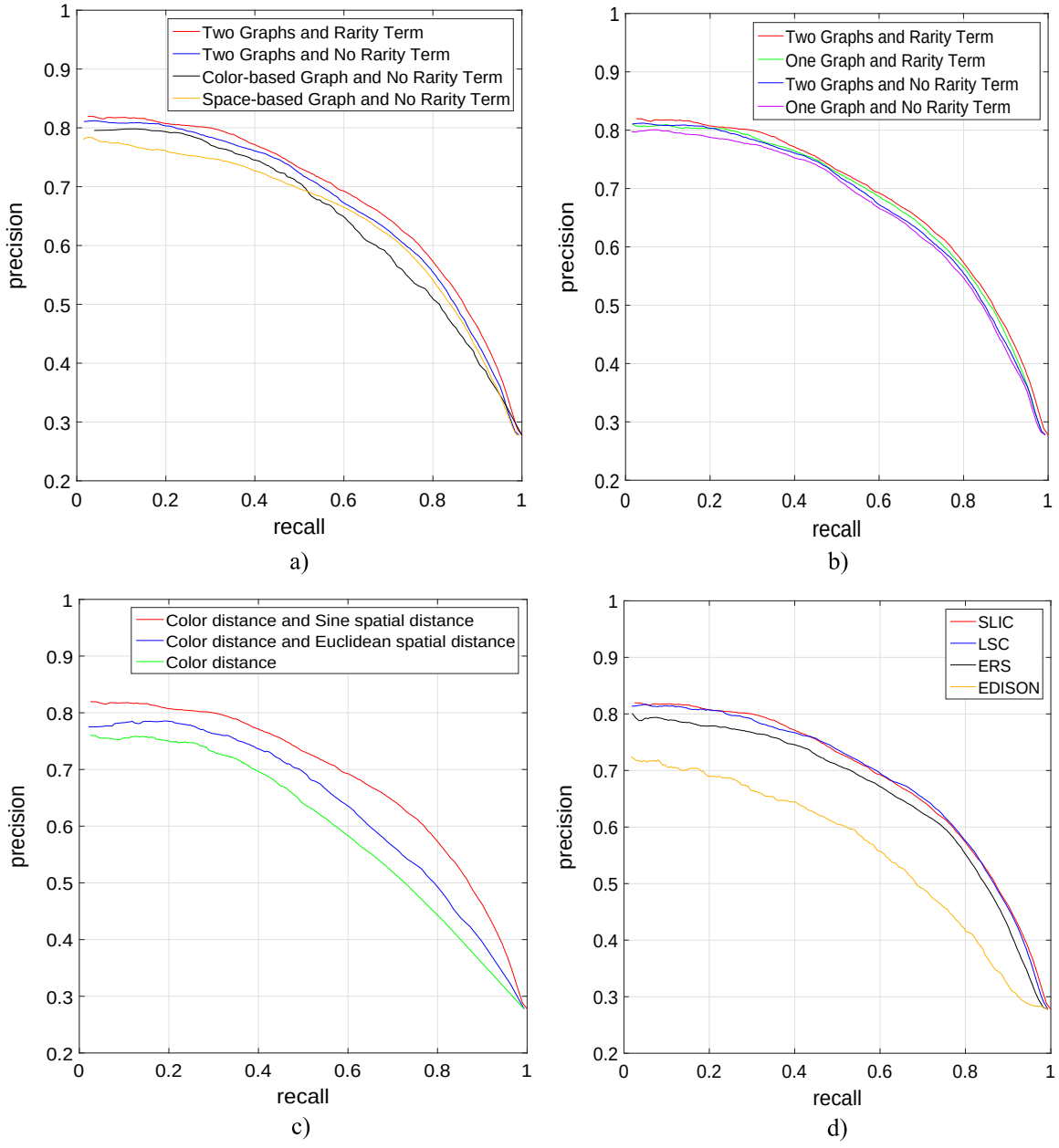

Fig. 4. Precision–recall curves on the SOD database with different design options of the proposed method. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

Table 5
Performance comparison with different weight for color distance.

ρ	0	0.2	0.4	0.6	**0.8**	**1**
AP	0.6377	0.6761	0.7080	0.7216	**0.7383**	**0.7368**
ρ	1.2	1.4	1.6	1.8	2	–
AP	0.7273	0.7109	0.6850	0.6602	0.6286	–

The optimal performance is highlighted with bold fonts.

nodes based on spatial locations and another graph is constructed based on color features) and without the rarity term. The above curves demonstrate that multiple graphs from different image properties are helpful to detect salient objects. The red curve represents the performance of the final output with two graphs and the rarity term. As shown by the blue curve and the red curve, the complete method outperforms the method without the rarity term, which demonstrates the effectiveness of our designed rarity term. Based on the above observations, both the use of multiple graphs based on different image properties and the rarity term in our optimization framework contribute to the overall performance.

We further compare the performances of the methods using two graphs and using one graph by connecting image nodes based on both spatial and color properties, shown in Fig. 4(b). No matter with or without the rarity term, the method using two graphs outperforms the method using one graph. One reason is that the method using two graphs would put emphasis on the spatial neighbors which share similar colors since both the graph based on spatial property and the graph based on color features would connect these close and similar nodes. For salient object detection problem, the spatial neighbors which share similar colors have a bigger chance to belong to the same class (foreground or background) than the spatial neighbors which do

Fig. 5. Visual comparison of different over-segmentation approaches. The column (a) is the original image and the remaining columns are results of different over-segmentation approaches.

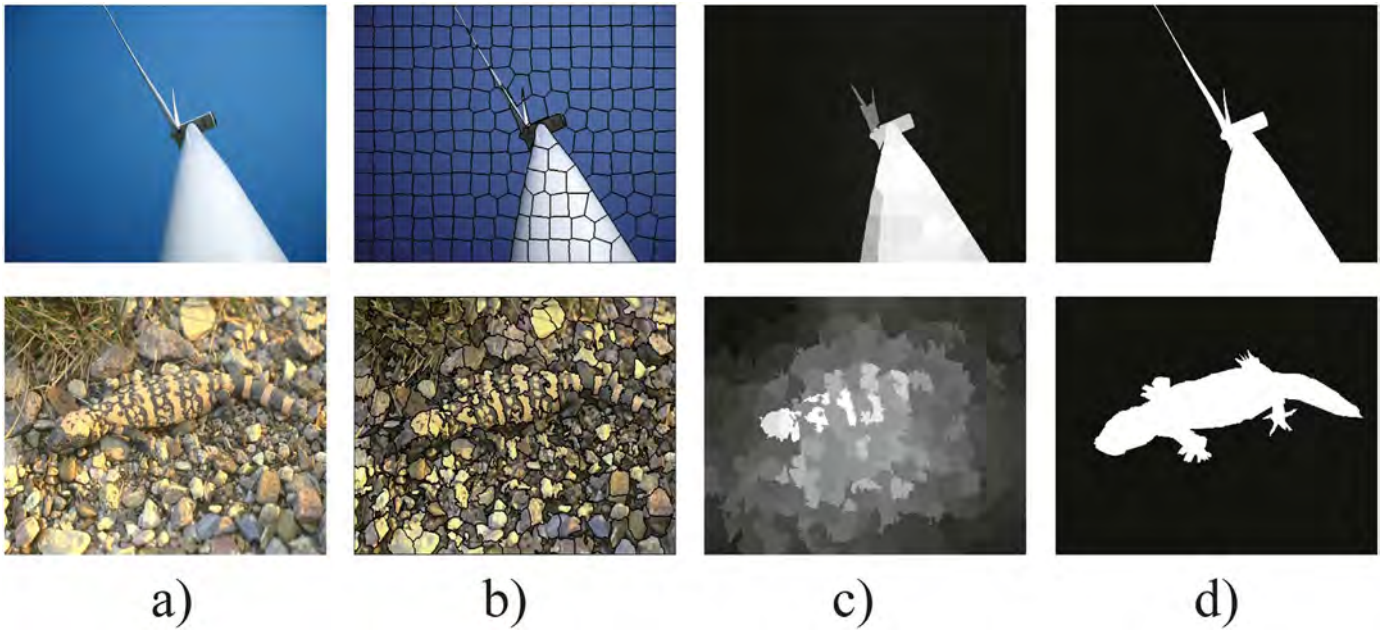

Fig. 6. Failed cases of our method. The column (a) is the original image; (b) contains the over-segmentation result; (c) is the imperfect saliency map; and (d) shows the corresponding truth.

not share similar colors and the similar image regions which are far from each other. Another reason is that in our method we use a weighted combination of the two graphs in the optimization framework. Weighted combination of two graphs is able to better describe the structure information contained in an image than one graph.

We also examine different edge weight computation methods using different distances in Fig. 4(c). In existing graph-based salient object detection methods, color distance is usually used to compute edge weight. The green curve in Fig. 4(c) provides the performance using color distance while the blue curve provides the performance using both color distance and Euclidean spatial distance. These two curves demonstrate that spatial distance also make contributions besides the color distance. The red curve provides the performance using both color distance and sine spatial distance, which proves that the sine spatial distance employed in our method is able to better detect salient objects in images than the frequently-used Euclidean spatial distance.

Since we use both color distance and Sine spatial distance for edge weight, we further examine the relative weights of color distance and Sine spatial distance to see whether one distance would dominate the overall distance value and thus have a dominant influence on the final saliency detection performance. The performances of the results using different weights ρ for color distance are shown in Table 5. The corresponding weight for Sine spatial distance is $2-\rho$. Then the final distance used for edge weight is $\rho \cdot D_c + (2-\rho) \cdot D_s$. We observe that the performance is poor when ρ equals 0 or 2 (i.e. only one distance is used) and the performance is best when ρ equals 0.8 or 1 (i.e. the color distance and sine spatial distance have nearly equal weight). This suggests that both distances make roughly equal contributions to the final saliency detection performance. In this paper, we simply sum these two distances (i.e. $\rho = 1$) as introduced in Eq. (6) for edge weight.

In this paper, we employ the commonly used segmentation approach: SLIC [43] to over-segment an image into homogeneous superpixels, as in most previous graph-based salient object detection methods [13,14,16,17]. We further compare four different over-segmentation approaches: EDISON [58], ERS [59], LSC [60] and SLIC [43] to examine their effects on the final saliency detection results. Visual comparison of these over-segmentation approaches is shown in Fig. 5. The superpixels segmented by EDISON have very different sizes, e.g. the sky in the image is mostly contained in one big superpixel while the crosses on the roof of the building in the second image is contained in small superpixels. Superpixels segmented by ERS are similar in size but have widely varying shapes. The superpixels segmented by LSC and SLIC tend to be similar in both size and shape. The quantitative comparison of the final saliency detection results

based on these approaches is shown in Fig. 4(d). The EDISON segmentation method gives the poorest saliency detection results. The ERS superpixel method outperforms EDISON, while the results based on LSC and SLIC give the best performance. From these comparisons, we observe that the variation in the sizes and shapes of the superpixels has an influence on superpixel-based salient object detection methods. Both LSC and SLIC give acceptable over-segmentation results for salient object detection. In this paper, we adopt the SLIC approach because it has been used in previous salient object detection methods.

The average execution time of the proposed method (without parallel programming) is 0.86 s per image conducted on a 64-bit PC with Intel Core i7-4790 CPU @ 3.60 GHz and 8 GB RAM.

5.6. Failed cases

Although our method achieves impressive results in most cases, it may fail if (1) the target is wrongly over-segmented by the superpixel method or (2) the target is extremely similar in appearance to the background. The two failed cases of our method are shown in Fig. 6. In Fig. 6(a), the rotor blades of the windmill are thin and long. Part of these rotor blades is wrongly over-segmented. In this case, our method can only detect part of the salient object. In Fig. 6(b), the appearance of the lizard is very similar to the background, therefore the final saliency map is not good. Actually, the above two cases (1) and (2) are also very challenging for the existing salient object detection algorithms.

6. Conclusion

In this work, we presented a novel graph-based optimization framework for salient object detection. Unlike the traditional graph-based optimization framework, we include multiple graphs and model the property of visual rarity as a regularization term in the optimization framework. The natural scene image is represented using multiple graphs representing different image properties over the set of its superpixels. The saliency value of each superpixel is computed based on the similarity between connected nodes in multiple graphs, saliency related queries, and visual rarity. An extensive experimental evaluation demonstrates the effectiveness and efficiency of our method on four representative saliency databases.

Acknowledgment

This work was supported by the National Natural Science Foundation (NNSF) of China (Grant numbers 61233011, 61374006, and 61473086); Major Program of National Natural Science Foundation (NNSF) of China (Grant number 11190015); Natural Science Foundation of Jiangsu (Grant number BK20131300); and the Innovation Fund of Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education (Grant number JYB201601).

References

[1] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (1998) 1254–1259.
[2] N. Bruce, J. Tsotsos, Saliency based on information maximization, in: Advances in Neural Information Processing Systems, 2005, pp. 155–162.
[3] M. Mancas, C. Mancas-Thillou, B. Gosselin, B. Macq, A rarity-based visual attention map –application to texture description, in: International Conference on Image Processing, 2006.
[4] T. Liu, J. Sun, N.-N. Zheng, X. Tang, H.-Y. Shum, Learning to detect a salient object, in: International Conference on Computer Vision and Pattern Recognition, 2007.
[5] X. Hou, L. Zhang, Saliency detection: a spectral residual approach, in: International Conference on Computer Vision and Pattern Recognition, 2007.
[6] L. Zhang, M.H. Tong, G.W. Cottrell, Sunday: saliency using natural statistics for dynamic analysis of scenes, in: Proceedings of the Thirty-first Annual Cognitive Science Society Conference, 2009.
[7] S. Goferman, L. Zelnik-Manor, A. Tal, Context-aware saliency detection, in: International Conference on Computer Vision and Pattern Recognition, 2010.
[8] Q. Zhao, C. Koch, Learning a saliency map using fixated locations in natural scenes, J. Vis. 11 (2011) 1–15.
[9] A. Borji, Boosting bottom-up and top-down visual features for saliency estimation, in: International Conference on Computer Vision and Pattern Recognition, 2012.
[10] N. Riche, M. Mancas, B. Gosselin, T. Dutoit, Rare: a new bottom-up saliency model, in: International Conference on Image Processing, 2012.
[11] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, N. Crook, Efficient salient region detection with soft image abstraction, in: International Conference on Computer Vision, 2013.
[12] L. Mai, Y. Niu, F. Liu, Saliency aggregation: a data-driven approach, in: Computer Vision and Pattern Recognition, 2013.
[13] C. Yang, L. Zhang, H. Lu, X. Ruan, M.-H. Yang, Saliency detection via graph-based manifold ranking, in: International Conference on Computer Vision and Pattern Recognition, 2013.
[14] S. Lu, V. Mahadevan, N. Vasconcelos, Learning optimal seeds for diffusion-based salient object detection, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2790–2797.
[15] J. Zhang, K. A. Ehinger, J. Ding, J. Yang, A prior-based graph for salient object detection, in: International Conference on Information Processing, 2014.
[16] C. Li, Y. Yuan, W. Cai, Y. Xia, D. Dagan Feng, Robust saliency detection via regularized random walks ranking, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2710–2717.
[17] C. Gong, D. Tao, W. Liu, S. J. Maybank, M. Fang, K. Fu, J. Yang, Saliency propagation from simple to difficult, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2531–2539.
[18] J. Zhang, J. Ding, J. Yang, Exploiting global rarity, local contrast and central bias for salient region learning, Neurocomputing 144 (2014) 569–580.
[19] A. Borji, L. Itti, State-of-the-art in visual attention modeling, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 185–207.
[20] A. Borji, D.N. Sihite, L. Itti, Salient object detection: a benchmark, in: European Conference on Computer Vision, 2012.
[21] H. Liu, I. Heynderickx, Studying the added value of visual attention in objective image quality metrics based on eye movement data, in: 2009 16th IEEE International Conference on Image Processing (ICIP), IEEE, Cairo, Egypt, 2009, pp. 1-5.
[22] A. Li, X. She, Q. Sun, Color image quality assessment combining saliency and FSIM, in: Fifth International Conference on Digital Image Processing, International Society for Optics and Photonics, 2013, pp. 88780I–88780I.
[23] J. Li, R. Ma, J. Ding, Saliency-seeded region merging: automatic object segmentation, in: Asian Conference on Pattern Recognition, 2011.
[24] C. Guo, L. Zhang, A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression, IEEE Trans. Image Process. 19 (1) (2010) 185–198.
[25] L. Itti, Automatic Foveation for video compression using a neurobiological model of visual attention, IEEE Trans. Image Process. 13 (10) (2004) 1304–1318.
[26] S. Goferman, A. Tal, L. Zelnik-Manor, Puzzle-like collage, in: Computer Graphics Forum, vol. 29, Wiley Online Library, 2010, pp. 459–468.
[27] J. Wang, L. Quan, J. Sun, X. Tang, H.-Y. Shum, Picture collage, in: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, IEEE, 2006, pp. 347–354.
[28] Y.-F. Ma, X.-S. Hua, L. Lu, H.-J. Zhan, A generic framework of user attention model and its application in video summarization, IEEE Trans. Multimed. 7 (5) (2005) 907–919.
[29] J. Ghosh, Y.J. Lee, K. Grauman, Discovering important people and objects for egocentric video summarization, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Providence, RI USA, 2012, pp. 1346–1353.
[30] F. Moosmann, D. Larlus, F. Jurie, Learning saliency maps for object categorization, in: International Workshop on the Representation and Use of Prior Knowledge in Vision (ECCV'06), 2006.
[31] H. Shen, S. Li, C. Zhu, H. Chang, J. Zhang, Moving object detection in aerial video based on spatiotemporal saliency, Chin. J. Aeronaut. 26 (5) (2013) 1211–1217.
[32] Z. Ren, S. Gao, L.-T. Chia, I.W.-H. Tsang, Region-based saliency detection and its application in object recognition, IEEE Trans. Circuits Syst. Video Technol. 24 (5) (2014) 769–779.
[33] S. Frintrop, M. Kessel, Most salient region tracking, in: 2009 IEEE International Conference on Robotics and Automation (ICRA'09), IEEE, Kobe, Japan, 2009, pp. 1869–1874.
[34] G. Zhang, Z. Yuan, N. Zheng, X. Sheng, T. Liu, Visual saliency based object tracking, in: Computer Vision (ACCV 2009), Xi'an, China, Springer, 2009, pp. 193–203.
[35] S. Wan, P. Jin, Y. Lihua, An approach for image retrieval based on visual saliency, in: International Conference on Image Analysis and Signal Processing, 2009.
[36] W. Zhu, S. Liang, Y. Wei, J. Sun, Saliency optimization from robust background detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2814–2821.
[37] B. Jiang, L. Zhang, H. Lu, C. Yang, M.-H. Yang, Saliency detection via absorbing Markov chain, in: IEEE International Conference on Computer Vision, 2013.
[38] N. Ahuja, M. Tuceryan, Extraction of early perceptual structure in dot patterns: integrating region, boundary, and component gestalt, Comput. Vis. Graph. Image Process. 48 (1989) 304–356.
[39] V. Gopalakrishnan, Y. Hu, D. Rajan, Random walks on graphs for salient object detection in images, IEEE Trans. Image Process. 19 (2010) 3232–3242.
[40] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in: International Conference on Computer Vision and Pattern Recognition, 2009, pp. 1597–1604.

[41] D. Grest, J.-M. Frahm, R. Koch, A color similarity measure for robust shadow removal in real-time, in: VMV, 2013.
[42] D. Zhou, O. Bousquet, T.N. Lal, J. Weston, B. Scholkopf, Learning with local and global consistency, in: IEEE Conference on Neural Information Processing Systems, 2003.
[43] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Susstrunk, SLIC superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2012) 2274–2282.
[44] B.W. Tatler, The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions, J. Vis. 7 (2007) 1–17.
[45] T. Judd, K. Ehinger, F. Durand, A. Torralba, Learning to predict where humans look, in: International Conference on Computer Vision, 2009.
[46] Y. Zhai, M. Shah, Visual attention detection in video sequences using spatiotemporal cues, in: ACM Multimedia, 2006.
[47] M.-M. Cheng, G.-X. Zhang, N.J. Mitra, X. Huang, S.-M. Hu, Global contrast based salient region detection, in: International Conference on Computer Vision and Pattern Recognition, 2011.
[48] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, Neural Inf. Process. Syst. (2006) 545–552.
[49] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, S. Li, Automatic salient object segmentation based on context and shape prior, in: British Machine Vision Conference, 2011.
[50] X. Li, Y. Li, C. Shen, A. Dick, A. van den Hengel, Contextual hypergraph modelling for salient object detection, in: IEEE International Conference on Computer Vision, 2013.
[51] Q. Yan, L. Xu, J. Shi, J. Jia, Hierarchical saliency detection, in: International Conference on Computer Vision and Pattern Recognition, 2013.
[52] P. Jiang, N. Vasconcelos, J. Peng, Generic promotion of diffusion-based salient object detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 217–225.
[53] D. Batra, A. Kowdle, D. Parikh, J. Luo, T. Chen, iCoseg: interactive co-segmentation with intelligent scribble guidance, in: Computer vision and Pattern Recognition, 2010.
[54] V. Movahedi, J.H. Elder, Design and perceptual validation of performance measure for salient object segmentation, in: IEEE Computer Society Workshop on Perceptual Organization in Computer Vision, 2010.
[55] D. Martin, C. Fowlkes, D. Tal, J. Malik, A data of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: IEEE International Conference on Computer Vision, 2001.
[56] M. Zhu, Recall, Precision and Average Precision, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo 2.
[57] G. Hripcsak, A.S. Rothschild, Agreement, the *f*-measure, and reliability in information retrieval, J. Am. Med. Inf. Assoc. 12 (3) (2005) 296–298.
[58] D. Comanicu, P. Meer, Mean shift: a robust approach toward feature space analysis, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 603–619.
[59] M.-Y. Liu, O. Tuzel, S. Ramalingam, R. Chellappa, Entropy rate superpixel segmentation, in: International Conference on Computer Vision and Pattern Recognition, 2011.
[60] Z. Li, J. Chen, Superpixel segmentation using linear spectral clustering, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Boston, MA USA, 2015, pp. 1356–1363.

Jinxia Zhang received the B.S. degree in the Department of Computer Science and Engineering, Nanjing University of Science and Technology, China in 2009 and the Ph.D. degree in the Department of Computer Science and Engineering, Nanjing University of Science and Technology, China in 2015. She was a visiting scholar in Visual Attention Lab, Brigham and Women's Hospital from 2012 to 2014. She is currently a lecturer in the School of Automation, Southeast University. Her research interests include visual attention, visual saliency detection, computer vision and machine learning.

Krista A. Ehinger received the B.S. degree in Engineering and Applied Science from the California Institute of Technology in 2003 and the B.Sc. (Hons) degree in Psychology from the University of Edinburgh in 2007. She received her Ph.D. degree in the Department of Brain and Cognitive Sciences at the Massachusetts Institute of Technology in 2013. She is currently a postdoctoral fellow in the Visual Attention Lab at Brigham and Women's Hospital and Harvard Medical School. Her research interests are in scene recognition, spatial representation, and visual search.

Haikun Wei received the B.S. degree in the Department of Automation, North China University of Technology, China in 1994, and the M.S. and Ph.D. degrees in the Research Institute of Automation, Southeast University, China in 1997 and 2000. He was a visiting scholar in RIKEN Brain Science Institute, Japan from 2005 to 2007. He is currently a professor in the School of Automation, Southeast University. His research interest is real and artificial in neural networks and industry automation.

Kanjian Zhang received the B.S. degree in Mathematics from Nankai University, China in 1994, and the M.S. and Ph.D. degrees in Control Theory and Control Engineering from Southeast University, China in 1997 and 2000. He is currently a professor in the School of Automation, Southeast University. His research is in nonlinear control theory and its applications, with particular interest in robust output feedback design and optimization control.

Jingyu Yang received his B.S. degree in Computer Science from Nanjing University of Science and Technology (NUST), Nanjing, China. From 1982 to 1984, he was a visiting scientist at the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign. From 1993 to 1994, he was a visiting professor in the Department of Computer Science, Missuria University. And in 1998 he acted as a visiting professor at Concordia University in Canada. He is currently a professor in the Department of Computer Science and Engineering at NUST. He is the author of more than 300 scientific papers in computer vision, pattern recognition and artificial intelligence. He has won more than 20 provincial and national awards. His current research interests are in the areas of pattern recognition, robot vision, image processing, data fusion and artificial intelligence.