Salient Region Detection via Integrating Diffusion-based Compactness and Local Contrast

Li Zhou, *Member, IEEE,* Zhaohui Yang, Qing Yuan, Zongtan Zhou, and Dewen Hu, *Senior Member, IEEE*

***Abstract*—Salient region detection is a challenging problem and an important topic in computer vision. It has a wide range of applications such as object recognition and segmentation. Many approaches have been proposed to detect salient regions using different visual cues such as compactness, uniqueness, and objectness. However, each visual cue-based method has its own limitations. After analyzing the advantages and limitations of different visual cues, we found that compactness and local contrast are complementary to each other. Additionally, local contrast can very effectively recover incorrectly suppressed salient regions using compactness cues. Motivated by this, we propose a bottom-up salient region detection method that integrates compactness and local contrast cues. Furthermore, to produce a pixel-accurate saliency map that more uniformly covers the salient objects, we propagate the saliency information using a diffusion process. Our experimental results on four benchmark datasets demonstrate the effectiveness of the proposed method. Our method produces more accurate saliency maps with better precision-recall curve and higher F-Measure than other nineteen state-of-the-arts approaches on ASD, CSSD, and ECSSD data sets.**

***Index Terms*—Salient region detection, compactness, local contrast, diffusion process, manifold ranking, random walks.**

I. Introduction

VISUAL attention is an important mechanism of the human visual system. It filters out redundant visual information and effectively selects highly relevant subjects, which are called the salient objects. Visual attention is considered to involve two mechanisms: stimulus driven [1] and task driven [2]. The stimulus-driven mechanism is often called bottom-up, and is fast, involuntary, and purely based on low-level visual stimuli. The task-driven mechanism is called top-down, and is based on high-level information such as prior knowledge of the task, emotions, and expectations. Accordingly, computational visual attention methods can be categorized into bottom-up [3-12] and top-down [13, 14] methods. In this paper, we focus on bottom-up salient region detection tasks.

Salient region detection methods aim to completely highlight entire objects of interest and sufficiently suppress background regions. Their output can be used for numerous computer vision problems such as image classification [15, 16], object detection and recognition [17, 18], image compression [19], and image segmentation [20, 21].

As a fundamental computer vision task, salient region detection has been extensively studied over the past few years, and a number of algorithms have been proposed [22-27]. Most bottom-up salient region detection methods rely on visual cues to consistently separate the salient object and background. These cues include uniqueness [3, 5, 7], compactness [9, 28, 29], and background [10, 12, 30].

Most uniqueness-based methods use low-level features of the image (such as intensity, color, and orientation) to determine the contrast between image pixels or regions and their surroundings. According to the contrastive reference regions, these methods can be roughly divided into local- and global contrast-based methods. Local contrast-based methods consider the uniqueness of pixels (or superpixels, image regions) with respect to their surrounding regions or local neighborhoods, whereas global contrast-based methods consider contrast relationships over the entire image.

Unlike uniqueness-based methods, which consider the uniqueness of low-level features in the feature space, compactness based methods consider the spatial variance of features. Ideally, salient pixels (or superpixels, image regions) tend to have a small spatial variance in the image space, whereas the background is distributed over the entire image and has a high spatial variance.

Background based methods use boundary and connectivity priors derived from common backgrounds in natural images [10]. These methods are primarily motivated by the psychophysical observations that salient objects seldom touch the image boundary, and most background regions can be easily connected to each other.

Although the above-mentioned methods have achieved good results in some aspects, each method has its own limitations. For example, Fig. 1 illustrates the saliency detection results using four state-of-the-art methods [5, 7, 9, 12]. Figure 1(c) shows that the local contrast-based method [5] tends to highlight the salient object's edges instead of uniformly propagating the saliency to the interior. The global contrast based method [7] sometimes produces high saliency values for non-salient regions, especially for regions with complex patterns or rare background distractors. This is shown in the first example of Fig. 1(d), where some grass regions in the background are highlighted. A typical limitation of the compactness based method [9] is that some salient regions may be wrongly suppressed when the foreground objects and background are similar. In the second example of Fig. 1(e), the inner smooth parts of the clock are wrongly suppressed.

This work was supported in part by the National Basic Research Program of China under Grant 2013CB329401 and in part by the National Natural Science Foundation of China under Grant 61375034.

Li Zhou, Zhaohui Yang, and Qing Yuan are with Naval Academy of Armament, Beijing, 100036, China. E-mail: lizhounaa@aliyun.com, zhyangnaa@sina.com, qingyuannaa@163.com

Zongtan Zhou and Dewen Hu are with the Department of Automatic Control, Collage of Mechatronics and Automation, National University of Defense Technology, Changsha, Hunan, 410073, P.R. China. E-mail: narcz@163.com, dwhu@nudt.edu.cn

Manuscript received October 20, 2014; revised

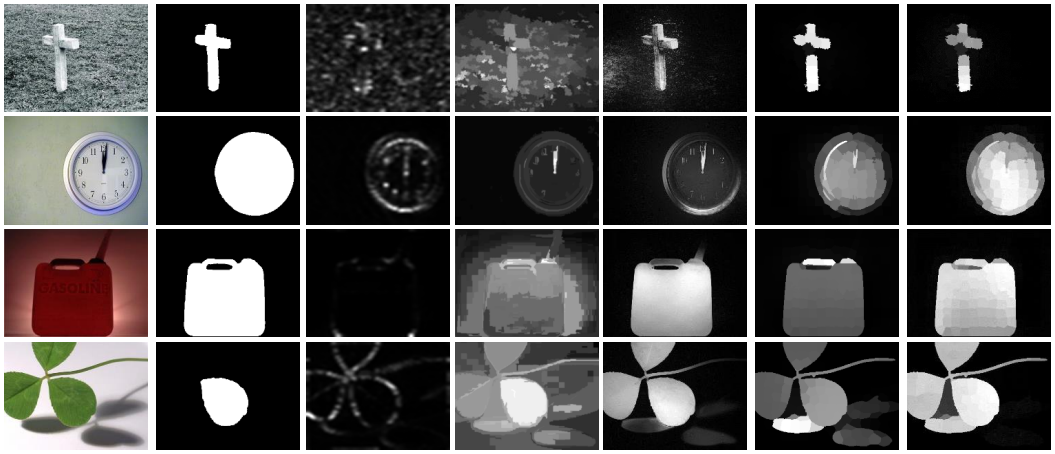

Fig. 1. Visual limitations of different methods. (a) Input image. (b) Ground truth salient regions. (c) Saliency maps using local contrast based method [5]. (d) Saliency maps using global contrast based method [7]. (e) Saliency maps using compactness based method [9]. (f) Saliency maps using background based method [12]. (g) Our method.

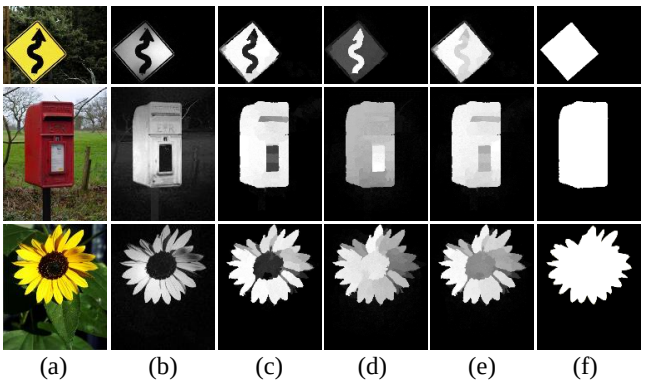

Fig. 2. Visual comparison of saliency maps produced by different cues. (a) Input image. (b) Saliency maps produced by compactness and global contrast integration [9]. (c) Saliency maps produced by our diffusion-based compactness method. (d) Saliency maps produced by our diffusion-based local contrast method. (e) Saliency maps produced by integrating compactness and local contrast (this work). (f) Ground truth.

Finally, background based methods [12] can perform very well. However, they fail when the salient objects touch the image boundary, as illustrated in the last two examples of Fig. 1(f).

From the above discussion, we can conclude that single visual cue based salient region detection methods all have their own limitations. To determine these limitations, different visual cues should be integrated into a unified framework. Motivated by this approach, some methods integrate multiple visual cues. Perazzi et al. [9] proposed a saliency filters method, which unifies uniqueness and compactness (of the spatial distribution) into a single, high-dimensional, Gaussian filtering framework. However, global contrast and compactness based methods have difficulty distinguishing between similar colors in the foreground and background. Consequently, the saliency filters method fails when foreground objects and the background are similar (e.g., the second example in Fig. 1(e)).

In this work, we integrated local contrast and compactness visual cues to generate saliency maps. Compared with the global contrast method, the local contrast method is a more appropriate complement to compactness. When the foreground is similar to some background regions, global contrast and compactness methods may wrongly suppress the foreground region. However, local contrast methods may properly highlight the foreground region based on the contrast with its neighboring region. This is shown in Fig. 2, which contains saliency maps from different visual cue-based methods (our compactness method, our local contrast method, an integrated compactness and global contrast method (saliency filters method) [9], and a compactness and local contrast integration method (this work)). We observed that the compactness and local contrast methods can complement each other. For example, in the first example of Fig. 2, the curve sign is wrongly suppressed by the compactness and saliency filters methods [9], whereas it is properly highlighted by our local contrast method. Furthermore, to mitigate the limitation of local contrast methods, which tend to highlight object boundaries rather than the entire area, we propagate the resulting saliency information of the local contrast using a diffusion process. We compared the performances of two different diffusion processes (random walks and manifold ranking), and found that manifold ranking produced the best results.

We evaluated the proposed approach on four datasets, and found that our method outperformed 19 existing state-of-the-art methods in three out of the four datasets.

The remainder of this paper is organized as follows. Section 2 contains a review of work related to bottom-up salient region detection. We describe diffusion processes for salient region detection in Section 3. In Section 4, we introduce our proposed salient region detection method. The experimental setup and results for the ASD [6], SOD [31], CSSD [24], and ECSSD [24] datasets are given in Section 5. Finally, Section 6 concludes the paper.

II. Related Work

Our work focuses on bottom-up salient region detection. A comprehensive survey of visual attention and saliency detection can be found in [32-34], and a quantitative analysis of different methods was provided in [35]. According to the type of visual cue, bottom-up salient region detection methods can be broadly classified into uniqueness, compactness, and background based. Furthermore, uniqueness-based methods can be roughly divided into local and global contrast-based techniques.

One of the first local contrast-based methods was the model of Itti et al. [3]. They used a difference of Gaussians approach to extract multi-scale color, intensity, and orientation information from images. This information was then used to define saliency by calculating center-surround d ifferences. Ma and Zhang [36] proposed an alternate local contrast analysis for generating saliency maps. They directly computed center-surround color differences in a fixed neighborhood for each pixel, and then extended the saliency map using a fuzzy growth model. Harel et al. [4] proposed a graph-based visual saliency method for non-linearly combining local uniqueness maps from different feature channels to concentrate conspicuity. Hou and Zhang [5] introduced a model in the frequency domain, which defines the saliency of a location based on the difference between the log-spectrum feature and its surrounding local average. Achanta et al. [37] calculated the saliency by computing center-surround contrasts of the average feature vectors, between the inner and outer sub-regions of a sliding square window. Liu et al. [38] computed center-surround histograms over windows of various sizes and aspect ratios in a sliding window. They trained a conditional random field to combine different features for salient object detection. Jiang et al. [8] used the difference between the color histogram of a region and its immediately neighboring regions to evaluate the saliency score.

Global contrast-based methods compute the saliency of individual pixels or image regions using contrast relationships over the complete image. Zhai and Shah [39] computed pixel-level saliency using the contrast with all other pixels. Bruce and Tsotsos [40] exploited Shannon's self-information measure in a local context to compute saliency. Achanta et al. [6] achieved globally consistent results based on a frequency-tuned method, which directly defines pixel saliency using the difference from the average image color. Goferman et al. [41] highlighted salient objects with their contexts by simultaneously modeling local low-level clues, global considerations, visual organization rules, and high-level features. Cheng et al. [7] proposed a regional contrast-based saliency extraction algorithm, which simultaneously considers the global region contrast over the entire image in the Lab color space and the spatial coherence, and used them to compute a saliency map. Lang et al. [42] detected salient positions by determining the consistently sparse elements from the entire image. Cheng et al. [43] proposed a soft image abstraction approach that captures large-scale perceptually homogeneous elements, thus effectively estimating global saliency cues. Zhu et al. [44] proposed a tag-saliency model for estimating the probability that each over-segmented region is salient, according to the global contrast of low- and high-level information in the scene.

Compactness-based methods have recently produced promising results. Gopalakrishnan et al. [28] considered that low-level features in the background have a larger spread than the salient regions. They presented a robust salient region detection framework based on the color and orientation distribution of images, and used the compact assumption to select a saliency map using the smaller spatial variances. Perazzi et al. [9] derived a pixel-accurate saliency map by simultaneously exploiting color and position to rate a region's uniqueness and spatial distribution. These are formulated in a unified way using high-dimensional Gaussian filters. Shi et al. [29] proposed a generic and fast computational framework called PISA, which imposes spatial prior terms on the color and structure contrast measures, so that the salient pixels are constrained to be compact and centered in the image. They concluded that fusing complementary contrast measures with a spatial prior significantly improved the effectiveness of the detection process. Cheng et al. [43] considered that a spatially compact distribution is another important saliency indicator, and is an important complementary cue to the contrast. They used the appearance similarity and spatial distribution of image pixels to produce perceptually accurate salient region detection results.

Unlike most of the previous salient region detection methods, Wei et al. [10] focused more on the background than the object, using background priors and geodesic distances for saliency detection. Since then, background-based methods have attracted considerable attention. Yang et al. [12] cast saliency detection into a graph-based ranking problem. It ranks the similarity of the image's elements with the background cues on a sparsely connected graph, to characterize the overall differences between the salient object and background. Li et al. [25] proposed a visual saliency detection algorithm from the perspective of reconstruction errors. They extracted image boundaries via superpixels, as likely cues for background templates, which they then use to construct dense and sparse appearance models. Jiang et al. [30] formulated saliency detection via an absorbing Markov chain on an image graph model, which uses the boundary superpixels around the image borders as the virtual background absorbing nodes.

There are many other cues such as "focusness" and "objectness". Chang et al. [45] presented a novel computational model to explore the relatedness of objectness and saliency. Jiang et al. [23] proposed a novel salient region detection algorithm by integrating three important visual cues: uniqueness, focusness, and objectness (UFO). Note that the above classification is not strict, because some research fits into many categories. For example, the saliency filters model of Perazzi et al. [9] can be classified into uniqueness based method or compactness based method. Zhu et al. [46] and Tong et al. [47] integrate local and global cues to generate saliency maps.

The approach most related to ours is Yang et al. [12], which casts saliency detection into a graph-based ranking problem. When ranking with background queries, they used nodes on the image boundary as background seeds. This may incorrectly suppress salient regions when the salient objects touch the

image boundary, as illustrated in the last two examples of Fig. 1(f). Unlike Yang et al. [12], we directly use foreground seeds obtained using the local contrast or compactness, to rank the saliency of the whole image regions.

III. Diffusion Processes

There has recently been a growing interest in using diffusion processes to propagate saliency information throughout a graph [4, 12, 30, 48, 49]. Harel et al. [4] used graph algorithms and a measure of dissimilarity to compute saliency in their graph-based visual saliency model. Yang et al. [12] cast saliency detection into a graph-based ranking problem. Jiang et al. [30] formulated saliency detection via an absorbing Markov chain on an image graph model. Gopalakrishnan et al. [48] used Markov random walks on two different graphs to detect the salient seed nodes. In [49], Lu et al. proposed a method for learning optimal seeds for object saliency using a diffusion process.

In all diffusion processes, the image is mapped into a graph $G = (V, E)$ with N nodes $\{v_1, v_2, \cdots, v_N\}$, and edges E weighted by an affinity matrix $\boldsymbol{W} = [w_{ij}]_{N\times N}$. Node v_i corresponds to the i^{th} image superpixel or patch and edge e_{ij} link nodes v_i and v_j to each other. After the graph is constructed, using a given vector of N saliency observations or saliency seeds $\boldsymbol{y} = [y_1, y_2, \cdots, y_N]^T$, diffusion process spread the seeds through the graph based on the defined affinity matrix. Finally we derive an object saliency map $\boldsymbol{f} = [f_1, f_2, \cdots, f_N]^T$.

In the following, we review the two most commonly used diffusion processes in salient region detection: random walk and manifold ranking. A detailed review of diffusion processes focusing on image retrieval was presented in [50].

A. Random Walk

Random walk is a popular diffusion process used in salient detection [4, 30 48]. The model interprets the diffusion processes as a random walk on the graph G, where a so-called transition matrix defines the probabilities for walking from one node to its linked nodes. Finally, the saliency is formulated as the equilibrium distribution of the random walk. The random walk transition matrix is defined as

$$\boldsymbol{P} = \boldsymbol{D}^{-1}\boldsymbol{W} \tag{1}$$

where $\boldsymbol{D} = diag\{d_{11}, d_{22}, \cdots, d_{NN}\}$, and d_{ii} is the degree of nodes v_i (i.e. $d_{ii} = \sum_j w_{ij}$). Then, a single step of the diffusion process is characterized by the simple iteration $\boldsymbol{f}^{t+1} = \boldsymbol{P}^T\boldsymbol{f}^t$. This standard random walk can be modified by introducing a random jump, such that at each step t, a random walk occures with probability α, while a random jump to an arbitrary node occures with probability $1-\alpha$. Thus, the diffusion process is

$$\boldsymbol{f}^{t+1} = \alpha\boldsymbol{P}^T\boldsymbol{f}^t + (1-\alpha)\boldsymbol{y} \tag{2}$$

where $1-\alpha$ is the random jump probability, and $\boldsymbol{y}$ are the saliency seeds that define the probabilities of randomly jumping to corresponding nodes. Following [49, 50], the iterated diffusion process can converge to an equilibrium distribution

$$\boldsymbol{f}^* = (1-\alpha)\left(\boldsymbol{I} - \alpha\boldsymbol{P}^T\right)^{-1}\boldsymbol{y} \tag{3}$$

where $\boldsymbol{I}$ is the identity matrix.

B. Manifold Ranking

Inspired by the standard PageRank approach, Zhou et al. [51] proposed a manifold ranking method that exploits the intrinsic manifold structure of data. The main procedure of this method is as follows. We first construct a graph on the data with an affinity matrix $\boldsymbol{W}$, and assign a ranking score to each query node (which is the saliency seed in this paper). Then, a diffusion process propagates the ranking score of each node to their nearby neighbors via the graph. The diffusion process repeats until a global stable state is achieved, and all nodes are ranked according to their final ranking scores. As illustrated in [12, 49], the optimal ranking is computed by minimizing an energy of the form

$$\sum\nolimits_{i,j=1}^{N} w_{ij}\left(\frac{f_i}{\sqrt{d_{ii}}} - \frac{f_j}{\sqrt{d_{jj}}}\right)^2 + \mu\sum\nolimits_{i=1}^{N}(f_i - y_i)^2 \tag{4}$$

where μ balances the smoothness constraint (the first term) and the fitting constraint (the second term). A single step of the diffusion process can be characterized by the iteration

$$\boldsymbol{f}^{t+1} = \alpha\boldsymbol{S}\boldsymbol{f}^t + (1-\alpha)\boldsymbol{y} \tag{5}$$

$$\boldsymbol{S} = \boldsymbol{D}^{-1/2}\boldsymbol{W}\boldsymbol{D}^{-1/2} \tag{6}$$

where $\alpha = 1/(1+\mu)$ specifies the relative contributions to the ranking scores from the neighbors and the initial ranking scores, and $\boldsymbol{S}$ is the normalized Laplacian matrix. By setting the derivative of (4) to zero, the closed form ranking solution can be written as

$$\boldsymbol{f}^* = (\boldsymbol{I} - \alpha\boldsymbol{S})^{-1}\boldsymbol{y} \tag{7}$$

Following [12], Equation (7) can be modified to another ranking function using the un-normalized Laplacian matrix

$$\boldsymbol{f}^* = (\boldsymbol{D} - \alpha\boldsymbol{W})^{-1}\boldsymbol{y} \tag{8}$$

We experimentally compared the saliency results using random walk and manifold ranking, and the results showed that manifold ranking outperforms the random walk. Hence, we use manifold ranking for the diffusion process in this work.

IV. Proposed Method

In this section, we present an efficient and effective saliency region detection method that integrates diffusion-based compactness and local contrast, as shown in Fig. 3. We first abstract the image into superpixels and construct a graph. Next, we compute two complementary saliency maps using the compactness visual cue and local contrast. The resulting saliency maps are propagated using a diffusion process and the constructed graph. Finally, we integrate the two computed saliency maps to generate a pixel-wise saliency map.

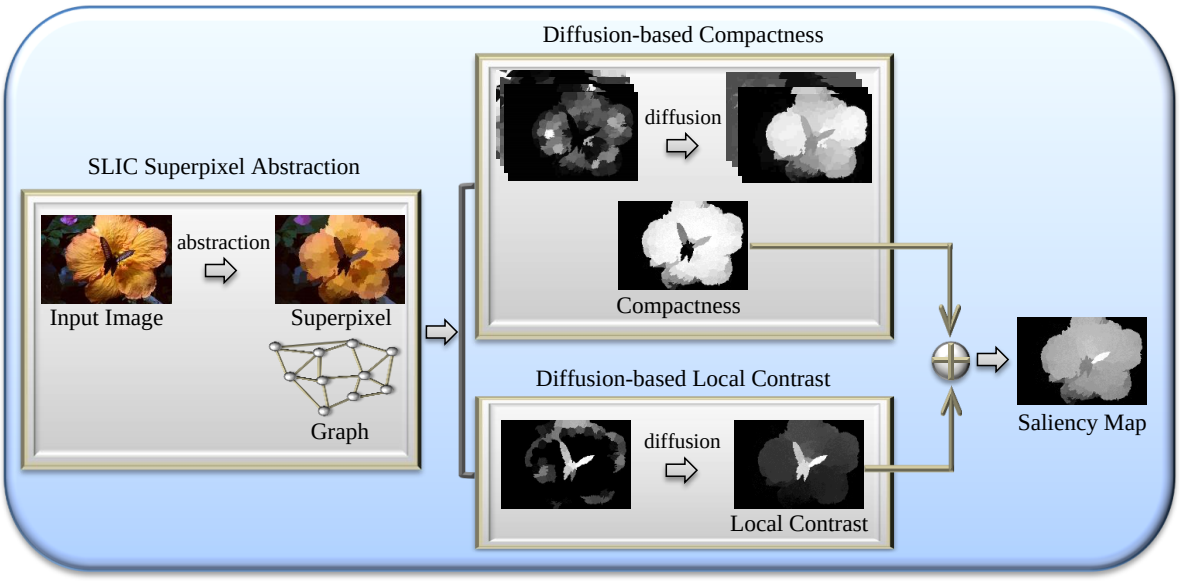

Fig. 3. Flowchart of the proposed method. It contains three stages: abstracting the image and construction the graph; detecting saliency using diffusion-based compactness and local contrast, and generating a pixel-wise saliency map that integrates compactness and local contrast based saliency maps.

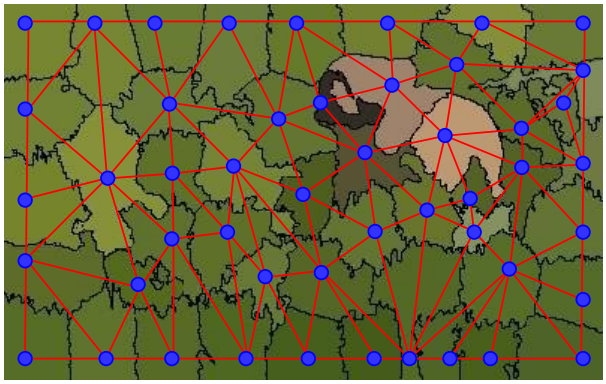

Fig. 4. Proposed graph model.

A. Graph Construction

Following the observation of Perazzi et al. [9] that abstracting an input image into homogeneous superpixels can improve the performance of salient object detection, we used the SLIC model [52] to abstract the input image into uniform and compact regions. After abstracting the image, we construct a graph $G = (V, E)$, as shown in Fig. 4. Each node corresponds to a superpixel generated by the SLIC model. Most existing algorithms [12, 30, 49] connect each node to neighboring nodes and nodes that share common boundaries with neighboring nodes (k-regular graph). However, in this graph, each node is only connected to its neighboring nodes. Additionally, each pair of boundary nodes are connected to each other to reduce the geodesic distance of similar superpixels.

In this work, we define the Lab color space distance l_{ij} between nodes v_i and v_j as

$$l_{ij} = \|\boldsymbol{c}_i - \boldsymbol{c}_j\| \tag{9}$$

where $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ are the mean of superpixels corresponding to nodes v_i and v_j in the Lab color space. Note that the distance matrix $\boldsymbol{L} = [l_{ij}]_{N\times N}$ is normalized to the interval [0, 1]. The affinity matrix $\boldsymbol{W}$ is defined as

$$w_{ij} = \begin{cases} e^{-\frac{l_{ij}}{\sigma^2}} & \text{if} j \in \mathcal{N}_i \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where σ is a constant, and $\mathcal{N}_i$ denotes the set of neighbors of node v_i. Note that all nodes around the image borders are considered neighbors of each other. Given the affinity matrix $\boldsymbol{W}$, the saliency propagation is implemented using Equation (8).

B. Diffusion-based Compactness

Salient objects generally correspond to real objects, therefore they are grouped together into connected regions. Therefore, salient objects typically have compact spatial distributions, whereas background regions have a wider distribution over the entire image. Motivated by this, we calculate the spatial variance of the superpixels.

We first define the similarity a_{ij} between a pair of superpixels v_i and v_j, using

$$a_{ij} = e^{-\frac{l_{ij}}{\sigma^2}} \tag{11}$$

To describe the similarity between superpixels more precisely, we propagate the similarity using the manifold ranking through the constructed graph. That is,

$$\boldsymbol{H}^T = (\boldsymbol{D} - \alpha\boldsymbol{W})^{-1}\boldsymbol{A} \tag{12}$$

where $\boldsymbol{A} = [a_{ij}]_{N\times N}$, and $\boldsymbol{H} = [h_{ij}]_{N\times N}$ is the similarity matrix after the diffusion process.

Salient objects are generally surrounded by background regions. Thus, in the spatial domain, the colors of background

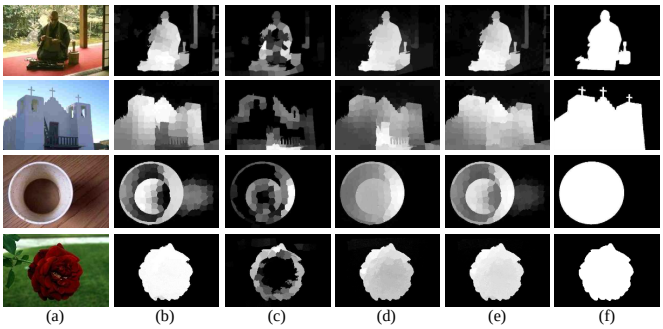

Fig. 5. Saliency maps with diffusion processes. (a) Original images. (b) Diffusion-based compactness. (c) Local contrast. (d) Diffusion-based local contrast. (e) Integrated results. (f) Ground truth

regions typically have a larger spread over the whole image, when compared with salient colors. Colors that exhibit large spatial variance across the image are less likely to be salient. We calculate the spatial variance of superpixel v_i using

$$\boldsymbol{sv}\left(i\right)=\frac{\sum_{j=1}^{N}h_{ij}\cdot n_j\cdot\left\|b_j-\mu_i\right\|}{\sum_{j=1}^{N}h_{ij}\cdot n_j} \tag{13}$$

where n_j is the number of pixels that belong to superpixel v_j, $b_j=\left[b_j^x,b_j^y\right]$ is the centroid of the superpixel v_j, and the spatial mean $\mu_i=[\mu_i^x,\mu_i^y]$ is defined as

$$\mu_i^x=\frac{\sum_{j=1}^{N}h_{ij}\cdot n_j\cdot b_j^x}{\sum_{j=1}^{N}h_{ij}\cdot n_j} \tag{14}$$

and

$$\mu_i^y=\frac{\sum_{j=1}^{N}h_{ij}\cdot n_j\cdot b_j^y}{\sum_{j=1}^{N}h_{ij}\cdot n_j} \tag{15}$$

A psychophysical study [53] showed that humans favor the center of images, and accordingly people usually frame objects of interest near the image center when taking photographs. Therefore, salient objects are more likely to be placed near the center of images, and the background generally spreads over the whole image. Motivated by this, we calculate the spatial distance of superpixels from the center of an image using

$$\boldsymbol{sd}\left(i\right)=\frac{\sum_{j=1}^{N}h_{ij}\cdot n_j\cdot\left\|b_j-p\right\|}{\sum_{j=1}^{N}h_{ij}\cdot n_j} \tag{16}$$

where $p=[p_x,p_y]$ is the spatial coordinate of the image center.

The preliminary saliency map calculated by diffusion-based compactness is finally defined using

$$S_{com}\left(i\right)=1-Norm\left(\boldsymbol{sv}\left(i\right)+\boldsymbol{sd}\left(i\right)\right) \tag{17}$$

where $Norm(x)$ is a function that normalizes x to the range between 0 and 1. The S_{com} saliency maps are shown in Fig. 5(b).

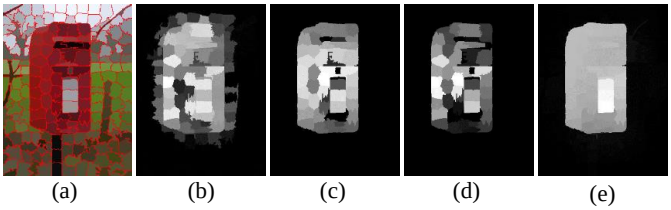

Fig. 6. Illustration of the main phases of our diffusion-based local contrast calculation. (a) Superpixels generated by SLIC. (b) The ***ld*** saliency map. (c) The ***dm*** saliency map. (d) The ***lc*** saliency map. (e) Saliency map produced by diffusion-based local contrast.

C. Diffusion-based Local Contrast

Although compactness based methods achieve good results in some aspects, they have limitations. When the foreground objects and background have similar appearances, some salient regions may be wrongly suppressed. To mitigate this, some approaches integrate the compactness visual cue with other cues. Perazzi et al. [9] unified the compactness and uniqueness into a single high-dimensional Gaussian filtering framework. However, we found that the local contrast is more complementary to compactness than the global contrast. When a foreground region is similar to some background regions, global contrast and compactness methods may wrongly suppress the foreground, whereas local contrast methods can highlight the foreground based on the contrast with its neighbor. In this section, we determine the saliency using the local contrast of an image superpixel with respect to its neighboring superpixels.

Although the local contrast can highlight the foreground regions that are wrongly suppressed by the compactness method, it may also highlight some background regions. Considering this, we use the saliency map calculated using compactness to suppress the incorrectly highlighted background regions. We define the Lab color space distance $\boldsymbol{ld}(i)$ of a superpixel v_i with respect to its neighborhood superpixels as

$$\boldsymbol{ld}\left(i\right)=\frac{\sum_{j\in\mathcal{N}_i}S_{com}\left(j\right)\cdot n_j}{\sum_{j\in\mathcal{N}_i}n_j}\cdot\max_{j\in\mathcal{N}_i}l_{ij} \tag{18}$$

A smaller $\boldsymbol{ld}(i)$ value corresponds to a higher probability that superpixel v_i belongs to the background. We set any value of ***ld*** that is less than the mean ***ld*** to zero. The ***ld*** saliency map is shown in Fig. 6(b). To enhance the reliability of the foreground detection (especially for complicated images) and improve the overall quality of salient region segmentation, we define the distribution measure for a superpixel v_i with respect to the centroid of saliency map S_{com} as

$$\boldsymbol{dm}\left(i\right)=\frac{\sum_{j=1}^{N}m_{ij}\cdot n_j\cdot\left\|b_j-\mu s\right\|}{\sum_{j=1}^{N}m_{ij}\cdot n_j} \tag{19}$$

where m_{ij} is the similarity between a pair of superpixels (v_i and v_j) according to their spatial distance , and $\mu s=[\mu s^x,\mu s^y]$ is the centroid of saliency map S_{com}. m_{ij} and μs are defined as

$$m_{ij}=a_{ij}\cdot e^{-\frac{\left\|b_i-b_j\right\|}{\sigma^2}} \tag{20}$$

$$\mu s^x = \frac{\sum_{i=1}^{N} S_{com}(i) \cdot n_i \cdot b_i^x}{\sum_{i=1}^{N} S_{com}(i) \cdot n_i} \tag{21}$$

and

$$\mu s^y = \frac{\sum_{i=1}^{N} S_{com}(i) \cdot n_i \cdot b_i^y}{\sum_{i=1}^{N} S_{com}(i) \cdot n_i} \tag{22}$$

The ***dm*** saliency map is shown in Fig. 6(c). We integrated the Lab color space distance (***ld***) and the distribution measure (***dm***) to determine the local contrast, that is,

$$\boldsymbol{lc}(i) = ld(i) \cdot (1 - Norm(\boldsymbol{dm}(i))) \tag{23}$$

where ***lc*** is a N dimensional column vector, The ***lc*** saliency map is shown in Fig. 5(c) and Fig. 6(d). We can see that the ***lc*** saliency map produced using the local contrast tends to highlight object boundaries rather than their entire area. We propagate the ***lc*** saliency map using manifold ranking as follow

$$S_{loc} = Norm((\boldsymbol{D} - \alpha \boldsymbol{W})^{-1} \boldsymbol{lc}) \tag{24}$$

where S_{loc} is a N dimensional column vector. The S_{loc} saliency maps are shown in Fig. 5(d) and Fig. 6(e). We can see that the S_{loc} saliency maps completely highlight the salient object regions.

D. Saliency Map Integration

The compactness and local contrast saliency cues methods efficiently produce two different saliency maps, S_{com} and S_{loc}. These maps are complementary to each other. We directly integrate these two different saliency maps to define the final saliency map,

$$S = Norm(S_{com} + S_{loc}) \tag{25}$$

The final saliency maps (*S*) are shown in Fig. 5(e). It can be seen from Fig. 2 and Fig. 5 that the compactness and local contrast saliency cues complement each other well.

V. Experimental Results

We ran several experiments to evaluate the proposed saliency region detection method using four public datasets.

1) ASD. The ASD dataset [6] contains 1,000 images selected from the MSRA database [38]. Each image is manually segmented into foreground and background. This dataset has been extensively used to test recently developed methods [9, 12, 25, 30].
2) SOD. The SOD dataset [31] is a collection of salient object boundaries based on the Berkeley segmentation dataset [54]. Seven subjects were asked to choose the salient object(s) in 300 images. The foreground mask in each image was provided by Wei et al. [10]. This dataset contains many images with multiple objects, which makes it a challenging test set.
3) CSSD. Yan et al. [24] constructed the complex scene saliency dataset (CSSD) using 200 images. They contain diversified patterns in both the foreground and background. Ground truth masks were produced by five subjects.
4) ECSSD. To represent natural image situations, Yan et al. [24] extended their CSSD dataset in to the larger ECSSD dataset, which contains 1000 images. It includes many semantically meaningful but structurally complex images. Ground truth masks were produced by five subjects.

Using These four data sets, we compared our method with 19 state-of-the-art approaches, including local contrast based approaches (IT [3], GB [4], SR [5], CSP [8], AC [37] and FG [36]), global contrast based approaches (FT [6], RC [7], MR [22], LC [39] and CA [41]), compactness based approach (SF [9]), background based approaches (GS [10], GBMR [12], DSR [25] and MC [30]), and multiple visual cues and integrated approaches (UFO [23], HS [24] and NDE [46]).

A. Experimental Setup and Evaluation Criterion

1) Experimental Setup: There are three parameters in the proposed method: N, the number of superpixel nodes used in the SLIC model; σ in Equations (10), (11), and (20), which controls the fall-off rate of the exponential function; and α in Equations (12) and (24), which balances the smooth and fitting constraints of the manifold ranking algorithm. For all four datasets, we experimentally set the parameters $N = 200$ and $\sigma^2 = 0.1$, and empirically set $\alpha = 0.99$ as in [51].

2) Evaluation Criterion: Similar to [6, 9, 12, 23, 30], we evaluated the performance of the salient region detection methods using two popular evaluation criterions: the average precision-recall curve and the F-Measure.

For the average precision-recall curve, we produced a binary mask of the salient object for a given saliency map using a threshold of $T_f \in [0, 255]$. To compare the quality of the different saliency maps, we varied this threshold from 0 to 255, and computed the precision and recall at each value of the threshold by comparing the binary mask and the ground truth. We then plotted a precision-recall curve using the sequence of precision-recall pairs. Combining the results from all the images of each dataset, we can obtain an average precision-recall curve.

We compute the F-Measure using

$$F_\beta = \frac{\left(1 + \beta^2\right) \cdot \text{Precision} \cdot \text{Recall}}{\beta^2 \cdot \text{Precision} + \text{Recall}} \tag{26}$$

Following [6], we set $\beta^2 = 0.3$, and applied an adaptive threshold T_a to the saliency map before calculating F_β. The threshold is defined as twice the mean saliency of the image, that is

$$T_a = \frac{2}{W \cdot H} \sum_{i=1}^{W} \sum_{j=1}^{H} S(i, j) \tag{27}$$

where W and H are the width and height of the saliency map S, respectively.

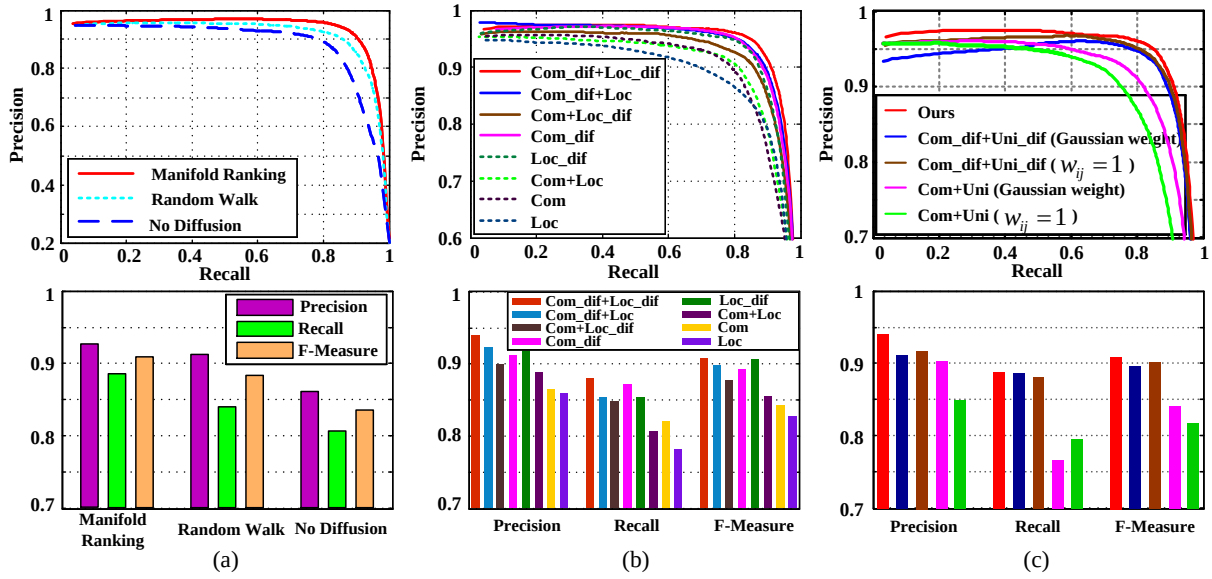

Fig. 7. Evaluation of the influence of various factors on saliency detection performance. Top: Precision-recall curves. Bottom: Precision, recall and F-Measure. (a) The influence of different diffusion processes. (b) The influence of different components. (c) The influence of integration. Com: compactness; Com_dif: diffusion based compactness; Loc: local contrast; Loc_dif: diffusion based local contrast; Uni: uniqueness; Uni_dif: diffusion based uniqueness.

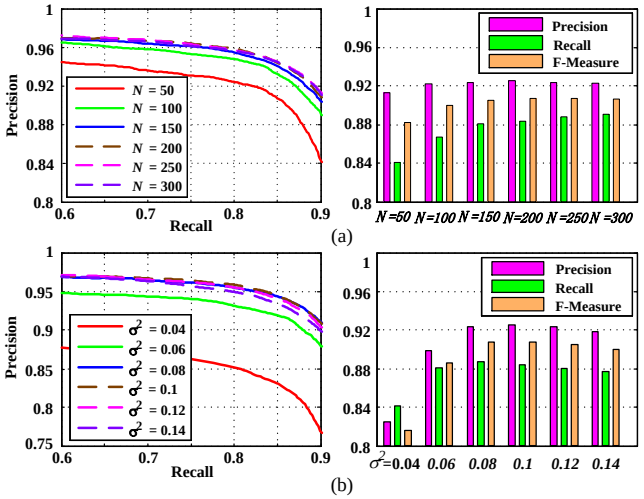

Fig. 8. Influence of N and σ on the graph. (a) Influence of N ($\sigma^2 = 0.1$). Left: precision-recall curves for different N. Right: precision, recall and F-Measure. (b) Influence of σ (N = 200). Left: precision-recall curves for different σ. Right: precision, recall and F-Measure.

B. Parameter Analysis

We carried out a series of experiments to investigate the influence of various factors on the saliency detection. These factors include the diffusion process algorithms, the integration of compactness and local contrast, the node connection mode on the graph, and the values of N and σ. The experiments in these investigations used the ASD data set.

1) Diffusion Process Algorithms: Our method uses diffusion processes to propagate the saliency information throughout the constructed graph. To validate it, we first investigated how the saliency detection performance is affected by the diffusion processes. We compared the performances using two diffusion process algorithms (random walk and manifold ranking) and without any diffusion process.

The precision-recall curves are plotted in the top of Fig. 7(a), the average precisions, recalls, and F-Measures using an adaptive threshold are shown in the bottom of Fig. 7(a). The performances using the diffusion processes are superior to not using any diffusion process. Additionally, the experimental results of Fig. 7(a) show that manifold ranking outperforms the random walk.

2) Components Analysis: We combined the two different visual cues of diffusion-based compactness and local contrast to obtain the final saliency map in this study. Thus, we experimentally evaluated the effectiveness of the two visual cue-based methods, and the effectiveness of the diffusion processes.

The precision-recall curves are plotted in the top of Fig. 7(b), and the average precisions, recalls, and F-Measures using an adaptive threshold are shown in the bottom of Fig. 7(b). These results demonstrate that the compactness and local contrast complement each other perfectly, and the diffusion processes improve the quality of both the compactness and local contrast based methods. Therefore, all the components of the saliency cues integration and diffusion processes contribute to the final results, and help to precisely detect the salient region.

3) Visual Cues Integration: In this paper, we show that the local contrast is a better complement to compactness than global contrast. We compared the performance of compactness combined with our local contrast method with the combination of local and global contrast proposed by Perazzi et al. [9], where the relation equation was $U_i = \sum_{j=1}^{N} \|\boldsymbol{c}_i - \boldsymbol{c}_j\|^2 \cdot w(b_i - b_j)$. A Gaussian weight w_{ij} yields a local contrast, whereas $w_{ij} = 1$ yields a global contrast.

The results including precision-recall curves, average precisions, recalls, and F-Measures are shown in Fig. 7(c). We

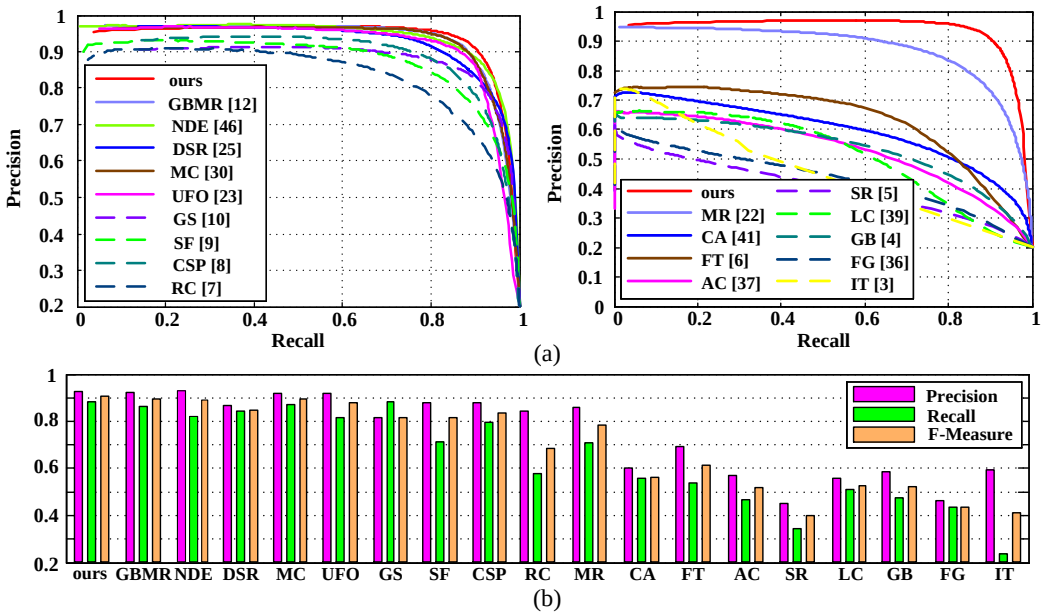

Fig. 9. Performance of the proposed method compared with 18 state-of-the-art methods, for the ASD database. (a) Average precision-recall curves for different approaches. (b) Average precision, recall, and F-Measure using different approaches with an adaptive threshold.

can see that the compactness combined with our local contrast method performs best. Figure 7(c) also demonstrates that the local contrast is a better complement to compactness than global contrast (In Fig. 7(c), Com+Uni with Gaussian weight compared with Com+Uni with $w_{ij} = 1$).

4) Parameters N and σ: N is a parameter used in the SLIC model to control the number of superpixel nodes, and σ is used to control the fall-off rate of the exponential function. The values of N and σ can have a significant effect on the performance. Thus, we experimentally examined the influence of these two parameters.

We used the SLIC model to abstract the input image into uniform and compact regions, where N controls the number of superpixel nodes. It is well known that N can have a significant effect on the performance. If N is too small, many different objects will be mapped to the same superpixel, which leads to a decrease in the detection of salient objects. If N is too large, salient objects will be mapped to many different superpixels, which may incorrectly suppress the salient regions. Thus, we first investigated how the saliency detection was affected by N. To determine an appropriate value for N, we experimented by letting it vary from 50 to 300 in intervals of 50.

The precision-recall curves and the average precisions, recalls, and F-Measures using an adaptive threshold are shown in Fig. 8(a), for different values of N. Figure 8(a) shows that, increasing N up to approximately 200 significantly improves the performance , but it makes little difference beyond this. Consequently, we used 200 as the optimal value for N in all subsequent experiments.

Figure 8(b) shows the performance variation for different values of σ. We can see that the performance increased gradually with σ, up to approximately $\sigma^2 = 0.1$. However, the performances decreased gradually as σ increased, when σ^2 was greater than 0.1. Consequently, we set $\sigma^2 = 0.1$ in all subsequent experiments.

C. Performance Comparison

We compared the performance of our method with 19 previously published results. We applied the comparative state-of-the-art methods to the ASD, SOD, CSSD, and ECSSD datasets, and compared the results using the average precision-recall curve and the F-Measure.

1) ASD data set: For the ASD data set, we compared our method with 18 state-of-the-art approaches (the well-known IT [3], GB [4], SR [5], FT [6], RC [7], CSP [8], SF [9], GS [10], MR [22], FG [36], AC [37], LC [39], CA [41] approaches, and four recently developed saliency models, GBMR [12], UFO [23], DSR [25], MC [30], and NDE [46]). These approaches were implemented using publicly available source code or we used the original published saliency detection results.

Figure 9 shows the evaluation results of the proposed method compared with the 18 state-of-the-art saliency detection approaches, for the ASD data set. The average precision-recall curves of Fig. 9(a) show that the proposed method performed better than the other approaches, for this dataset. Note that the proposed method outperformed the GBMR [12] and MC [30] methods, which are also based on a diffusion process.

The average precision, recall, and F-Measure using different approaches with an adaptive threshold are shown in Fig. 9(b). The proposed method achieved the highest F-Measure values. The two evaluation criteria consistently show that the proposed method outperformed all the other approaches. The CA [41], CSP [8], and RC [7] were the top-performing approaches for saliency detection in a recent benchmark study [35], and GBMR [12], UFO [23], DSR [25], and MC [30] are more recently developed, state-of-the-art approaches. Some

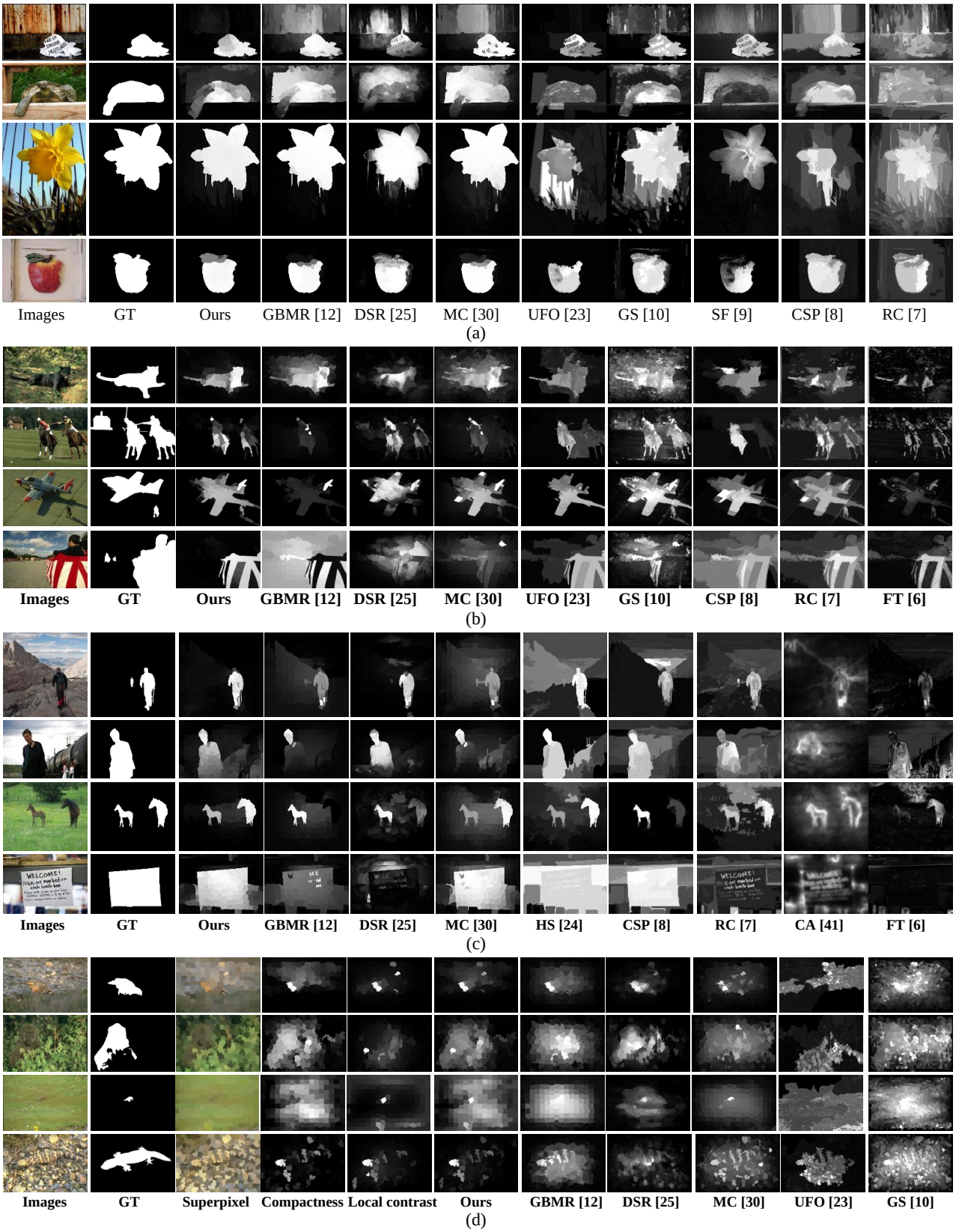

Fig. 10. Our visual image saliency maps compared to recently developed, state-of-the-art approaches. (a) The ASD data set. (b) The SOD data set. (c) The ECSSD data set. (d) Failure cases on the SOD data set.

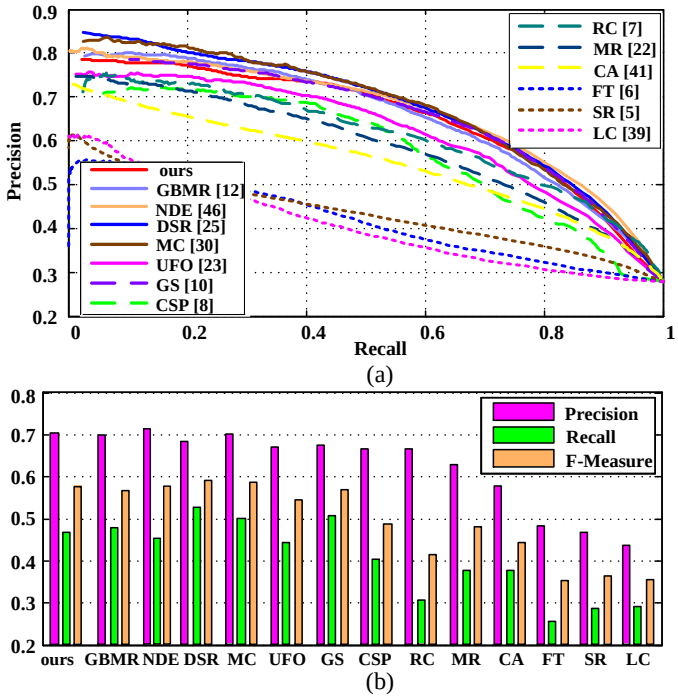

Fig. 11. Performance of the proposed method compared with 13 state-of-the-art methods, for the SOD database. (a) Average precision-recall curves for different approaches. (b) Average precision, recall, and F-Measure using different approaches with an adaptive threshold.

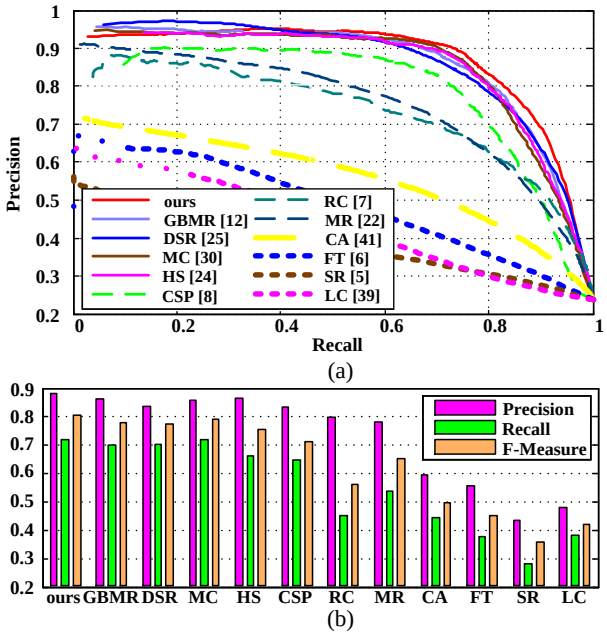

Fig. 12. Comparison of the proposed method with 11 state-of-the-art methods, for the CSSD database. (a) Average precision-recall curves for different approaches. (b) Average precision, recall, and F-Measure using different approaches with an adaptive threshold.

visual comparison examples are shown in Fig. 10(a), and more results can be found in the supplementary material. Note that the proposed method more completely highlights the salient regions with well-defined boundaries while adequately suppressing the backgrounds, when compared with the other methods.

2) SOD data set: Using the SOD data set, we qualitatively and quantitatively compared the proposed method with 13 state-of-the-art approaches (SR [5], FT [6], RC [7], CSP [8], GS [10], GBMR [12], MR [22], UFO [23], DSR [25], MC [30], LC [39], CA [41] and NDE [46]).

Figure 11(a) shows the average precision-recall curves, and Fig. 11(b) shows the average precision, recall, and F-Measure with an adaptive threshold, for all 13 approaches. Figure 11 shows that the proposed method performed as well as the GS and GBMR approaches, and achieved a slightly worse precision-recall curve and lower F-Measure than the DSR and MC approaches. However, note that the proposed method outperformed all the other approaches, including RC, CSP, and CA which were the top-performing methods for saliency detection in a recent benchmark study [35].

Fig. 10(b) shows some visual comparison examples. The proposed method effectively highlighted the salient object regions with well-defined boundaries, and suppressed the background regions.

3) CSSD data set: Using the CSSD data set, we qualitatively and quantitatively compare the proposed method with 11 state-of-the-art approaches, including SR [5], FT [6], RC [7], CSP [8], GBMR [12], MR [22], HS [24], DSR [25], MC

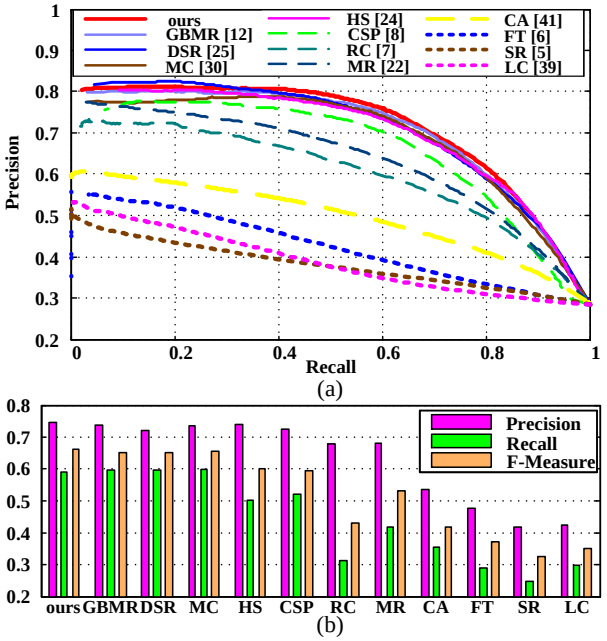

Fig. 13. Comparison of the proposed method with 11 state-of-the-art methods, for the ECSSD database. (a) Average precision-recall curves for different approaches. (b) Average precision, recall, and F-Measure using different approaches with an adaptive threshold.

[30], LC [39] and CA [41].

Figure 12(a) compares our precision-recall curve with the other 11 state-of-the-art approaches. As shown in Fig. 12(a), the proposed method performed better than the other approaches for recall values from 0.4 to 1, but it performed poorly compared with the GBMR and DSR approaches for recall values from 0 to 0.4.

The average precision, recall, and F-Measure using different approaches with an adaptive threshold are shown in Fig. 12(b). Our method achieved the highest precision and F-Measure values, and its recall measure was comparable to MC. Note that the original HS work of Yan et al. [24] achieved significantly worse precision-recall curve, precision, recall, and F-Measure values than our method.

4) ECSSD data set: To represent typical natural images, Yan et al. [24] extended their CSSD dataset into a larger dataset (ECSSD) of 1000 images. Using the ECSSD dataset, we qualitatively and quantitatively compared the proposed method with 11 state-of-the-art approaches, including SR [5], FT [6], RC [7], CSP [8], GBMR [12], MR [22], HS [24], DSR [25], MC [30], LC [39], and CA [41].

Figure 13(a) shows the average precision-recall curves of all 11 saliency detection approaches. The proposed method performed better than the other approaches for recall values from 0.3 to 1, but it performed poorly compared with the DSR approaches for recall values from 0 to 0.3. Figure 13(b) shows the average precision, recall, and F-measure with an adaptive threshold. The proposed method performed as well or slightly better than the GBMR, DSR, and MC approaches in terms of the F-Measure, and outperformed the other approaches. Some visual comparison examples are shown in Fig. 10(c). The proposed method more completely highlighted the salient regions with well-defined boundaries while adequately suppressing the backgrounds, when compared with the other methods.

D. Failure Cases

Figure. 10(d) shows some failures generated by our method and recently developed state-of-the-art approaches, for the SOD dataset. As previously stated, the proposed method only uses the color feature. That is, the local contrast uses the color contrast in the color space between superpixels, whereas compactness uses the spatial variance of the color in the image space. The proposed method mainly depends on color information, therefore it may fail for images that do not have much color variation, especially when foreground and background objects have similar colors. Figure 10(d) shows that the estimated salient object is not salient according to its color contrast or compactness, as shown in the corresponding maps in Fig. 10(d). This limitation could be tackled by incorporating more features such as shape, texture, or even high-level knowledge. Note that, as shown in Fig. 10(d), the other recently developed state-of-the-art approaches also did not perform well in these cases.

VI. Conclusion

In this paper, we proposed a bottom-up method for detecting salient regions in images by integrating two complementary visual cues (compactness and local contrast) with diffusion processes. After considering the advantages and limitations of different visual cues, we found that compactness and local contrast are complementary to each other. Additionally, local contrast can effectively recover the incorrectly suppressed salient regions using compactness cues. To produce a pixel-accurate saliency map that more uniformly covers the salient objects, we propagate the saliency information using a manifold ranking diffusion process on a graph. Our experimental results using four benchmark datasets demonstrated the effectiveness of the proposed method; it produced more accurate saliency maps with better precision-recall curves and higher F-measures than 19 state-of-the-art approaches, when applied to the ASD, CSSD, and ECSSD datasets.

References

[1] H. E. Egeth and S. Yantis, "Visual attention: control, representation, and time course," *Ann. Rev. Psychologogy*, vol. 48, pp. 269-297, 1997.

[2] J. M. Henderson, "Human gaze control during real-world scene perception," *Trends in Cognitive Sciences*, vol. 7, pp. 498-504, 2003.

[3] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 20, no. 11, pp. 1254-1259, 1998.

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

[5] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2007, pp. 1-8.

[6] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2009, pp. 1597-1604.

[7] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2011, pp. 1063-6919.

[8] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li. "Automatic salient object segmentation based on context and shape prior," in *Proc. Brit. Mach. Vis. Conf*., 2011, pp. 1-12.

[9] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2012, pp. 733-740.

[10] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Proc. Eur. Conf. Comput. Vis*., 2012, pp. 29-42.

[11] Y. Xie, H. Lu, and M.-H. Yang, "Bayesian saliency via low and mid level cues," *IEEE Trans. Image Process*., vol. 22, no. 5, pp.1689-1698, 2013.

[12] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2013, pp. 3166-3173.

[13] C. Kanan, M. H. Tong, L. Zhang, and G. W. Cottrell, "Sun: Topdown saliency using natural statistics," *Visual Cognit*., vol. 17, no. 8, pp. 979-1003, 2009.

[14] J. Yang and M.-H. Yang, "Top-down visual saliency via joint crf and dictionary learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2012, pp. 2296-2303.

[15] C. Siagian and L. Itti, "Rapid biologically-inspired scene classification using features shared with visual attention," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 29, no. 2, pp. 300-312, Feb. 2007.

[16] G. Sharma, F. Jurie, and C. Schmid, "Discriminative spatial saliency for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2012, pp. 3506-3513.

[17] D. Walthera, U. Rutishausera, C. Kocha, and P. Peronaa, "Selective visual attention enables learning and recognition of multiple objects in cluttered scenes," *Comput. Vis. Image Understand*, vol. 100, pp. 41-63, 2005.

[18] B. Alexe, T. Deselaers, and V. Ferrari, "Measuring the objectness of image windows," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 34, no. 11, pp. 2189-2202, 2012.

[19] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Trans. Image Process*., vol. 19, no. 1, pp. 185-198, Jan. 2010.

[20] L. Wang, J. Xue, N. Zheng, and G. Hua, "Automatic salient object extraction with contextual cue," in *Proc. IEEE Int. Conf. Comput. Vis*., 2011, pp. 105-112.
[21] C. Jung and C. Kim, "A unified spectral-domain approach for saliency detection and its application to automatic object segmentation," *IEEE Trans. Image Process*., Vol. 21, pp. 1272-1283, 2012.
[22] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2012, pp. 853-860.
[23] P. Jiang, H. Ling, J. Yu, and J. Peng, "Salient region detection by UFO: uniqueness, focusness and objectness," in *Proc. IEEE Int. Conf. Comput. Vis*., 2013, pp. 1976-1983.
[24] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2013, pp. 1155-1162.
[25] X.-H. Li, H.-C. Lu, L.-H. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in *Proc. IEEE Int. Conf. Comput. Vis*., 2013, pp. 2976-2983.
[26] R.-S. Liu, J.-J Cao, Z.-C. Lin, and S.-G. Shan, "Adaptive partial differential equation learning for visual saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2014, pp. 3866-3873.
[27] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu, "Saliency detection on light field," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2014, pp. 2806-2813.
[28] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Salient region detection by modeling distributions of color and orientation," *IEEE Trans. on Multimedia*, vol. 11, no. 5, pp. 892-905, 2009.
[29] K.-Y. Shi, K.-Z. Wang, J.-B. Lu, and L. Lin, "PISA: Pixelwise image saliency by aggregating complementary appearance contrast measures with spatial priors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2013, pp. 2115-2122.
[30] B.-W. Jiang, L.-H. Zhang, H.-C. Lu, C. Yang, and M.-H Yang, "Saliency detection via absorb Markov chain," in *Proc. IEEE Int. Conf. Comput. Vis*., 2013, pp. 1665-1672.
[31] V. Movahedi and J.H. Elder, "Design and perceptual validation of performance measures for salient object segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2010, pp. 49-56.
[32] A. Toet, "Computational versus psychophysical bottom-up image saliency: A comparative evaluation study," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 33, no. 11, pp. 2131-2146, 2011.
[33] A. Borji, D. Sihite, and L. Itti, "Quantitative analysis of human model agreement in visual saliency modeling: A comparative study," *IEEE Trans. Image Process*., vol. 22, no. 1, pp. 55-69, 2013.
[34] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 35, no. 1, pp. 185-207, 2013.
[35] A. Borji, D. N. Sihite, and L. Itti, "Salient object detection: A benchmark," in *Proc. Eur. Conf. Comput. Vis*., 2012, pp. 414-429.
[36] Y.-F. Ma and H. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *ACM Multimedia*, 2003, pp. 374-381.
[37] R. Achanta, F.J. Estrada, P. Wils, and S. Süsstrunk, "Salient region detection and segmentation," in *ICVS*, 2008, pp. 66-75.
[38] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 33, pp. 353-367, 2011.
[39] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *ACM Int'l Conf. Multimedia*, 2006, pp. 815-824.
[40] N. Bruce and J. Tsotsos, "Saliency based on information maximization," in *Advances in Neural Information Processing Systems*, 2006, pp. 155-162.
[41] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2010, pp. 2376-2383.
[42] C. Lang, G. Liu, J. Yu, and S. Yan, "Saliency detection by multitask sparsity pursuit," *IEEE Trans. Image Process*., vol. 21, pp. 1327-1338, 2012.
[43] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zhang, V. Vineet, and N. Crook, "Efficient salient region detection with soft image abstraction," *in Proc. IEEE Int. Conf. Comput. Vis*., 2013, pp. 1529-1536.
[44] G.-K. Zhu, Q. Wang, and Y. Yuan, "Tag-Saliency: Combining bottom-up and top-down information for saliency detection," *Comput. Vis. Image Underst*., vol. 118, pp. 40-49, 2014.
[45] K.-Y. Chang, T.-L. Liu, H.-T. Chen, and S.-H. Lai, "Fusing generic objectness and visual saliency for salient object detection," in *Proc. IEEE Int. Conf. Comput. Vis*., 2011, pp. 914-921.
[46] L. Zhu, D.A. Klein, S. Frintrop, Z.-G. Cao, and A.B. Cremers, "A Multisize Superpixel Approach for Salient Object Detection Based on Multivariate Normal Distribution Estimation," *IEEE Trans. Image Process*., vol.23, no.12, pp. 5094-5107, Dec. 2014.
[47] N. Tong, H.-C. Lu, Y. Zhang, and X. Ruan, "Salient object detection via global and local cues," *Pattern Recognition*, 2015, http://dx.doi.org/10.1016/j.patcog.2014.12.005
[48] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Random walks on graphs for salient object detection in images," *IEEE Trans. Image Process*., vol. 19, no. 12, pp. 3232-3242, Dec. 2010.
[49] S. Lu, V. Mahadevan, and N. Vasconcelos, "Learning optimal seeds for diffusion-based salient object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2014, pp. 2790-2797.
[50] M. Donoser and H. Bischof, "Diffusion processes for retrieval revisited," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2013, pp. 1320-1327.
[51] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf, "Ranking on data manifolds," in *Advances in Neural Information Processing Systems*, 2004, pp. 169-176.
[52] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 34, no. 11, pp. 2274-2282, 2012.
[53] B. Tatler, "The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions," *Journal of Vision*, vol. 7, no. 14, pp. 1-17, 2007.
[54] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. IEEE Int. Conf. Comput. Vis*., 2001, pp. 416-423.

Li Zhou received the B.S. and M.S. degrees from Dalian Navy Academy, China, in 2004 and 2006, respectively. He got his Ph.D. degree from the National University of Defense Technology in 2012. From 2012, he was with the Naval Academy of Armament. His research interests include computer/biological vision, visual navigation, and machine learning.

Zhaohui Yang received the B.S. and M.S. degrees from PLA University of Science and Technology, China, in 1995 and 1998, respectively. From 1998, she was with the Naval Academy of Armament. His research interests include visualization and machine learning.

Qing Yuan received the B.S. and M.S. degrees from Naval university of engineering, China, in 1999 and 2002, respectively. From 2002, he was with the Naval Academy of Armament. His research interests include machine learning and data mining.

Zongtan Zhou received the B.S., M.S. and Ph.D. degrees from the National University of Defense Technology, China, in 1990, 1994 and 1998, respectively. From February 2010 to February 2011, He was a Visiting Scholar with the Eberhard Karls Universitt Tübingen. He was promoted Professor in 2007. His research interests include image/signal processing, computer/biological vision, neural networks, cognitive neuroscience and brain-computer interface.

Dewen Hu received the B.S. and M.S. degrees from Xi'an Jiaotong University, China, in 1983 and 1986, respectively. From 1986, he was with the National University of Defense Technology. From October 1995 to October 1996, he was a Visiting Scholar with the University of Sheffield, UK. He got his Ph.D. degree from the National University of Defense Technology in 1999. He was promoted Professor in 1996. His research interests include image processing, system identification and control, neural networks, and cognitive science. He is an action editor of Neural Networks.