A Novel Saliency Detection Method via Manifold Ranking and Compactness Prior

Libo Zhang, Zakir Ullah, Yihan Sun, and Tiejian Luo
University of Chinese Academy of Sciences
Beijing, China

***Abstract*—For improving the performance of saliency detection, several algorithms used graph construction have achieved excellent results. This paper proposes a novel bottom-up approach of saliency detection, which takes the advantages of both prior background and compactness. At first, we optimize the image boundary selections, by removing erroneous sections with a fixed threshold, to achieve more accurate saliency estimation results. The saliency map obtained by ranking with background queries can be optimized with compactness prior. The objects of salient are connected regions which are group together, with a compact form which are spatial distributed. Compared to the 8 state-of-the-art saliency detection approaches, our experimental results which test on the three public datasets show that the proposed algorithm improves accuracy and robustness significantly. This algorithm can find its potential applications in many different areas, but it is best suit for medical science and technologies because of high accuracy requirements. It can be used in the medical imaging processing to accurately differentiate tumor from bones, muscles and fats.**

I. Introduction

Selective mechanism of visual attention is a critical mechanism of the human visual system, which makes human the capability to locate the most salient parts of a scene [1]. It effectively selects the relevant visual information of a scene to remove redundancy. Similar to the function of the selective mechanism, saliency detection plays the role for identifying the most important and informative part of a scene or an object. Visual attention consists of two mechanisms: stimulus and task driven [2]. The first one is stimulus-driven mechanism [3], also called bottom-up, which is involuntary, fast, and totally based on low-level visual stimuli. On the other hand, the task-driven mechanism is often called top-down, which is based on high-level information such as previous knowledge of the task, expectations and emotions. All computational visual attention methods can be categorized into bottom-up or top-down methods according to the different mechanisms present.

Salient region detection is still a challenging task because the human attention system has not been fully understood yet and has considered as a major research area in recent years. Many methods has been proposed by the researchers around the world. Among them object recognition and detection [4], [5], image retrieval [6], [7], image and video compression [8], [9], as well as image segmentation [10], [11] are some typical applications of saliency detection. Duo to high accuracy and robustness of our algorithm, it can find its potential applications in bioinformatics, medical sciences and medical image processing for tumor detection [12].

We mainly focus on the bottom-up salient region detection tasks in this work. Firstly, we obtain saliency maps by ranking with background queries. Many previous saliency detection approaches are based on manifold ranking, in which, the four sides of image are defined as background. Unlike those methods, here boundary influences are optimized by locating and removing the erroneous regions that are most different from the other three boundaries. Next we optimize the saliency map obtained by ranking with background queries and compactness prior as the salient objects generally are grouped together into connected regions. The visual comparison of the proposed method is shown in figure 1.

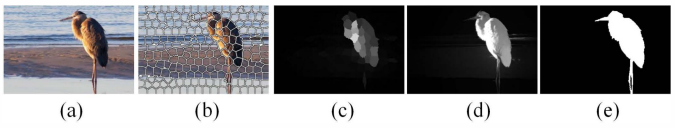

Fig. 1. Visual comparison of the proposed method. (a) Input image. (b) SLIC segmentation result. (c) Saliency map. (d) Full resolution saliency map. (e) Ground truth.

The proposed method is examined on the three datasets. From the experiment results, we observe that our method achieves better performance than 8 existing state-of-the-art methods. The structure of this paper is organized as follows: we have a review of related work about bottom-up and top-down salient region detection approaches in Section 2. In Section 3, we present the proposed method for salient region detection. We test our method on three datasets in Section 4. Finally, Section 5 gives the conclusion.

II. Related Work

Our work is a bottom-up mechanism for salient region detection. According to its domain's, bottom-up salient region detection mechanisms can be classified into frequency domain and spatial or spatiotemporal domain based. Most of bottom-up methods are spatiotemporal domain based, which can be broadly categorized as compactness, uniqueness and background based [13].

In the early times, one of the saliency detection method was proposed by Itti et al. [14]. They presented a visual attention system which used a difference of Gaussians approach to extract multi-scale intensity, color and orientation features from an image. Then the multi-scale image features were combined

978-1-5090-1610-5/16/$31.00 ©2016 IEEE

into a single topographical saliency map by calculating center-surround differences. Similar to [14], J. Harel et al. [15] first extracted feature vectors to form an activation map. Then they defined Markov chains of the different maps, and then treated the equilibrium distribution of the map location as activation and saliency values. Liu et al. [16] formulated salient region detection as an image segmentation problem, which separated the salient object from the image background. They extracted center-surround histogram, multi-scale contrast and color spatial distribution to describe a salient object locally, regionally and globally. Then they trained a random conditional field for combining various features to detect salient object. They also defined a large image dataset contained thousands of labeled images as MSRA and widely used it. Zhai and Shah [17] used luminance feature to define pixel-level saliency, based on a pixel's contrast to all other pixels. Cheng et al. [18] proposed a regional contrast on basis of saliency detection method that could evaluate global contrast differences and spatial coherence. Stas Goferman et al. [19] proposed a context-aware saliency detection method. Contrary to many other methods which only detected salient regions, they aimed to detect the image regions that contained the scene, so the special part of background regions might be marked as significant area. Cheng et al. [20] suggested a soft image abstraction methodology that could capture large-scale elements which were perceptually homogeneous. Gopalakrishnan et al. [21] proposed a robust salient region detection framework, which was based on orientation and color distribution of an image, he applied the compact assumption for selection of a saliency map by using the smaller spatial variances. Ren et al. [22] presented a region based saliency detection algorithm and applied it to object recognition. They are based on an adaptive mean shift algorithm for extracting super-pixels and measuring saliency of each super-pixel through its spatial compactness.

There are also some excellent frequency domain based saliency detection method. Xiaodi Hou et al. [23] analyzed the logarithmic spectrum of input image to obtain the spectral residual which was then transformed into spatial domain. The saliency map was obtained by using Inverse Fourier Transform. Based on the theory in [23], Chenlei Guo et al. [24] proposed that phase spectrum Fourier transform played a key role in detecting the salient regions. This method was successfully applied into image and video compression. On the base of above two methods, Li et al. [25] proposed a saliency detection method that used the convolution of the image amplitude spectrum with a low-pass Gaussian kernel of a suitable scale. Achanta et al. [11] used the DoG filter for band pass filtering of image and calculated pixel saliency via using difference of the average image color in CIELAB color space. Li et al. [26] found that the visual saliency possibly hid in the phases of middle frequencies, and defined the saliency detector by the assistance of prior knowledge obtained through supervised and unsupervised machine learning.

Some top-down based or combining of bottom-up and top-down methods also achieved excellent results. Itti and Koch [27] first put forward the theory that applied top-down factors into saliency detection model for the first time, then they found there was a definite relation between visual attention and eye movements. D. Gao et al. [28] introduced a one-versus-all classification problem, which was a top-down saliency detection mechanism. Jimei Yang et al. [29] formulated the top-down model based on joint learning of a Conditianal Random Field (CRF) and a discriminative dictionary. Ali Borji [30] combined low-level features which included intensity, orientation, color and saliency maps of prior best bottom-up models with top-down cognitive visual features (e.g., humans, faces and cars, etc.) and learned a direct mapping from those features to eye fixations by using SVM, Regression and AdaBoost classifiers. Tian et al. [31] used the features of color and orientation to calculate bottom-up saliency map and fused top-down cue of depth-from-focus from a single image.Zhu et al. [32] introduced a tag-saliency model for estimating the probability, that each over-segmented region is salient.

III. The Proposed Algorithm

In this section, we present an effective and efficient mechanism for saliency region detection, which uses ranking with background queries and compactness prior. In order to provide more accurate saliency estimations, we first optimize the image boundary selection via removing erroneous regions by a fixed threshold. The saliency map obtained by ranking with background queries can be optimized with compactness prior. Salient objects generally are group together into connected regions, which have compact spatial distributions.

A. Manifold Ranking

Manifold and Random walk ranking are two widely used diffusion processes for detection of salient region. Li Zhou et al. [13] compared the experimental results of saliency detection by using random walk and manifold ranking, the comparison showed that manifold ranking outperformed the random walk. Thus, we used manifold ranking for the diffusion process in this research work.

Manifold ranking is firstly proposed in [33], it exploits the intrinsic structure of the manifolds of data (such as image) for graph labeling, which is inspired from the standard Page-Rank algorithm. Given a dataset , $X = \{x_1, \cdots, x_q, x_{q+1}, \cdots x_n\} \in R^{m \times n}$, the first q elements are labeled queries and the rest q_n elements are ranked according to their relevance to the queries.Let $f : X \rightarrow R^n$ denotes a ranking function for assigns a ranking value f_i to each point x_i , and f can be considered as a vector $f = [f_1, f_2, \cdots, f_n]^T$. Let $y = [y_1, y_2, \cdots, y_n]^T$ denotes an indication vector, then $y_i = 1$, if x_i is a query, otherwise $y_i = 0$. Next we establish a graphical structure $G = (V, E)$ with edge E and node V, where V belongs to the dataset X and edges E are weighted by an affinity matrix $W = [w_{ij}]_{n \times n}$. $D = diag\{d_{11}, \cdots d_{nn}\}$ in the degree matrix, where $d_{ii} = \sum_j w_{ij}$.The following

minimization problem is used to calculate the optimal ranking of queries:

$$f^* = \arg\min_f \frac{1}{2}\left(\sum_{i,j=1}^{n}\left\|\frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_j}}\right\|^2 + \mu\sum_{i=1}^{n}\|f_i - y_i\|^2\right) \tag{1}$$

μ controls balance of the smoothness constraint (the first term) and the fitting constraint (the second term). The smoothness constraint limits a ranking function not changing too much between nearby points and the fitting constraint limits the ranking function not differing too much from the initial query. The above function can be represented by a single iteration:

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

where α=$1/(1+\mu)$ is the relative contributions to the ranking scores from the neighbors and the initial ranking score, and $S = D^{-1/2}WD^{-1/2}$ is the normalized Laplacian matrix. If we set the derivative of function (1) equals to zero, the closed form ranking function can be written as follow:

$$f^* = (I - \alpha S)^{-1}y \tag{3}$$

In function (3), I denotes an identity matrix. Manifold ranking can be considered as a one-class classification problem [34], where only negative or positive examples are required. The above equation can be modified to another ranking function by using the non-normalized Laplacian matrix:

$$f^* = (D - \alpha W)^{-1}y \tag{4}$$

From the conclusion of [35], the latter equation can achieve a better performance. Hence, equation (4) is used in this work.

B. Graph Construction

We design a graphical representation, $G = (V, E)$, where E is a set of undirected edges and V is a set of nodes. We use SLIC algorithm [36] for abstracting the input image into compact and uniform regions. Each node corresponds to a super pixel generated through the SLIC algorithm. Therefore, each node is connected to its neighboring nodes, because the neighboring nodes are most likely to share similar saliency values and appearance. Furthermore, each pair of nodes at boundary regions is connected to each other to reduce the geodesic distance of similar super pixels, thereby improving the ranking results [35].

The designed graph has a sparse connection. Hence, most elements of the affinity matrix are zero value. In above graphical representation, the weight between each pair of nodes are defined as follow:

$$w_{ij} = \begin{cases} e^{-\frac{\|c_i - c_j\|}{\sigma^2}} & j \in \mathrm{N}_i \\ 0 & otherwise \end{cases} \tag{5}$$

where c_i and c_j are the mean of superpixels corresponding to nodes v_i and v_j respectively in the Lab color space, and σ controlling parameters for the strength of weight. N_i denotes the set of neighbors of node v_i. According to the ranking of nodes on the designed graph, the inverse matrix $(D - \alpha W)^{-1}$ in Eq. (4) can be considered as a complete affinity matrix, and the saliency propagation is implemented by using Equation (4).

C. Background Prior

Many manifold ranking based saliency detection method [13], [35] defined the regions on the four sides of image as background. However, it is a possibility for the boundary regions of the input image to be occupied by the foreground object [37]. Using such a boundary for background saliency estimation of queries may lead to undesirable results, especially when the foreground object is adjacent to the boundary. Therefore, we minimize the boundary influences by identifying and removing erroneous boundaries before the process of background saliency estimation, and hence can improve manifold ranking based method.

It's conspicuous to distinguish the difference between salient region and background region. Similarly, the boundary that is adjacent to foreground tends to have different color distribution compared to the last three boundaries. To select the erroneous boundary, we treat the superpixel regions of the adjacent boundaries as connected areas, furthermore we calculate their normalized pixel-wise RGB histogram as follows:

$$H_O(h) = \frac{1}{n}\sum_{i=1}^{n}\delta(I_i - h) \tag{6}$$

where $O \in \{top, bottom, right, left\}$ represents the locations of four boundaries; n is the pixel number of superpixel region; h is the intensity bin variable and the value interval of it is 0 to 255; I_i is the intensity value of pixel I; $\delta()$ is the unit impulse function,

$$\delta(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{7}$$

The Euclidean distance between any two of the four histograms is calculated as follows,

$$E(b_1, b_2) = \sqrt{\sum_{h=0}^{255}[(H_{b_1}^R(h) - H_{b_2}^R(h))^2 + (H_{b_1}^G(h) - H_{b_2}^G(h))^2 + (H_{b_1}^B(h) - H_{b_2}^B(h))^2]} \tag{8}$$

In above formula R, G, B refers to red, green, and blue channel respectively. By calculating all values of Euclidean distance can obtain a 4×4 symmetric matrix E, which is denoted as, the column whose column sum is maximal corresponds to the boundary that is most different from three other boundaries. Then we normalize the four column sums of matrix E. If the maximum column sum $S_{\max}$ and minimum column sum $S_{\min}$ after normalization satisfy $S_{\max} - S_{\min} \geq 0.4$, the boundary correspond to the maximum column sum is recognized as erroneous and should be removed. If $S_{\max} - S_{\min} < 0.4$, the four boundaries are retained and all the regions adjacent to boundaries are recognized as background regions.

D. Ranking with Background Queries

For the remaining three sides of the image, the superpixels on each side are labeled as background seeds, which are

elements of the indication vector y in Eq.2, and other nodes are labeled as zeros. Then the labeled nodes (query samples) are used to rank the relevance to all the other regions. We construct three (or four) saliency maps by using boundary priors and then put them together for the final map. According to Eq.3, if one boundary is removed, three ranking results will be achieved afterwards, and the subscript "bou" corresponds to the remaining three boundaries.

As an example considering the top boundary, some nodes on it are labeled as queries and the reaming nodes as unlabeled data. Thus, the indicator vector y is given, nodes are ranked, according to Eq. 3, based on N dimensional vector, where N is the total number of nodes in graph. Each element in this vector(y) indicates the correspondence of a node to the background queries, and its complement is the irrelevance to the background queries, which is saliency measure. The vector is normalized to the range [0,1], and the saliency map using the top boundary prior can be written as follow:

$$S_t(i) = 1 - \overline{f_t}^*(i) \qquad i = 1, \cdots, N \tag{9}$$

Where i represents a superpixel node of graph, N is the total superpixel number, and $\overline{f_t}^*$ is the normalized vector.

Similarly, the other saliency maps using the bottom, left or right image boundary as queries can be computed. Then the saliency map that rank with background queries can be calculated by the following process:

$$S_b(i) = \prod_{bou} S_{bou}(i), \qquad i = 1, \cdots, N \tag{10}$$

Advantage of removing the erroneous boundary is the improvement in the accuracy of results, in the cases where one or more boundaries are adjacent to the foreground object. As Figure.1 shown, the saliency map that removes the erroneous boundary (right for the first image, and bottom for the second image) is more accurate than the saliency map that labels four boundaries as background queries.

E. Compactness Prior

The saliency map obtained by ranking with background queries can be optimized with compactness prior. Salient objects are generally group together into connected regions with a typical compact spatial distribution, whereas this this background regions of the image have a wider distribution.

Firstly, we define the similarity a_{ij} between two superpixels v_i and v_j as follow:

$$a_{ij} = e^{-\frac{\|c_i - c_j\|}{\sigma^2}} \tag{11}$$

Then by means of the manifold ranking, the similarity between superpixels is propagated through the designed graph. From Eq. 4, the similarity matrix is calculated by using the formula:

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

Where $A = [a_{ij}]_{N \times N}$, and $H = [h_{ij}]_{N \times N}$ is the similarity matrix. Then we compute the saliency distribution centroid of saliency map S_b:

$$\overline{P_b} = \frac{\sum_{i=1}^{N} P_i \cdot n_i \cdot S_b(i)}{\sum_{i=1}^{N} n_i \cdot S_b(i)} \tag{13}$$

Where n_i is the number of pixels that correspond to the superpixel (v_i), P_i is position coordinate of superpixel v_i, $S_b(i)$ is the saliency value of v_i which is normalized to range [0,1]. The superpixels near the saliency distribution centroid have high saliency values, and vice versa. Inspired from this, we calculate the spatial distance of superpixels from the saliency distribution centroid by using

$$sd(i) = \frac{\sum_{j=1}^{N} h_{ij} \cdot n_j \cdot \left\| P_j - \overline{P_b} \right\|}{\sum_{j=1}^{N} h_{ij} \cdot n_j} \tag{14}$$

The above spatial distance is normalized to range between 0 to 1. Then the saliency map calculated by compactness prior is defined as follow:

$$S_{cp}(i) = 1 - sd(i) \tag{15}$$

F. Final Saliency Map

Final saliency map is calculated by the integration of two saliency maps as follow:

$$S(i) = S_b(i) \cdot S_{cp}(i) \tag{16}$$

Integration of the maps is followed an up-sampling method in [38] to get full resolution saliency map.

IV. Experiment Results

Due to evaluating the performance of the proposed model, we test our approach with 8 state-of-the-art methods on MSRA10K and ECSSD datasets, and verify the design options on ASD dataset.

A. Experimental Setup and Evaluation Criterion

The proposed method used in this research work has three parameters: N, the number of superpixel nodes used in the SLIC algorithm; σ in Equations (5) and (11), which controls the fall-off rate of the exponential function; α in Equations (2), (3), (4) and (12), which balances the smoothness and fitting constraints of the manifold ranking algorithm. In this experiments, we choose $N = 200$, σ=0.1, and α=0.99.

Evaluation criterion: Similar to many saliency detection methods, we use average precision, recall and F-measure to evaluate the performance of different saliency detection methods. We calculated the precision rate and recall rate as follows:

$$precision = \frac{TP}{TP + FP}, recall = \frac{TP}{TP + FN} \tag{17}$$

Where TP (true positive) represents the number of pixels inside the salient regions of saliency map and ground truth map. Whereas, FP (false positive) is the number of pixels inside the salient regions of saliency map and non-salient

regions of ground truth map. FN is the number of pixels inside the salient regions of ground truth map and non-salient regions of saliency map. We compute the F-measure by following formula:

$$F_\beta = \frac{(1+\beta^2)\cdot precision \cdot recall}{\beta^2 \cdot precision + recall} \tag{18}$$

Similar to [38], we set $\beta^2 = 0.3$ to give more weight to the precision than recall. An adaptive threshold (T_a) is applied to segment the saliency map before calculating. Where T_a is two times the mean saliency of a given image:

$$T_a = \frac{2}{W \times H} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} S(x,y) \tag{19}$$

W and H in Eq. (19) are the width and height of the saliency map respectively. $S(x,y)$ is the saliency value of a pixel at position (x,y).

B. Examination of Design Options

At first we inspect the novelty of our proposed algorithm on the ASD dataset, and the result is given in Figure 2. In the results the green and blue curves show the final saliency output comparison with and without the erroneous boundary removal respectively. For improving the curve of proposed, the erroneous boundary is removed. Next, we build the saliency maps without compactness prior. The final saliency map also excels the saliency map without using compactness prior, as shown by the blue and red curves in Fig 2.

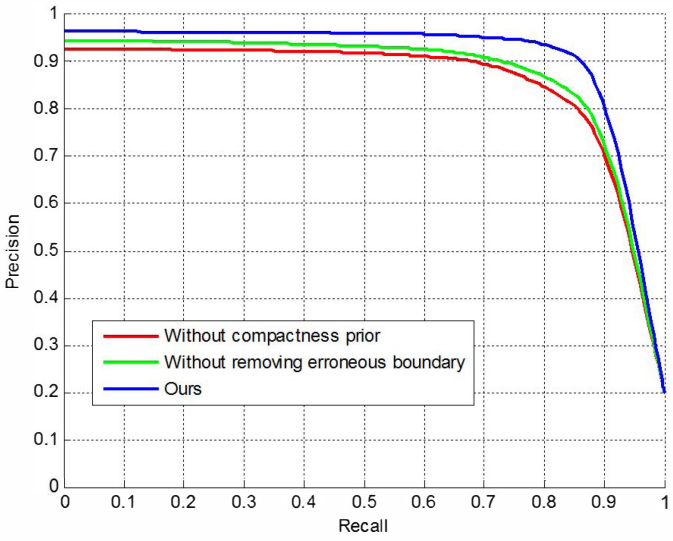

Fig. 2. Precision-recall curves on the ASD dataset with different design options of the proposed approach.

Based on above observations, both the erroneous boundary removal and the compactness prior contribute to the overall performance. Therefore, we adopt both of them in the following evaluations.

C. Comparison with State-of-the-art methods

We compare our proposed algorithm for this research with 8 state-of-the-art saliency detection methodologies, including GMR[35], HS[39], SF[38], FT[11], CLC[13], RC[18], CA[19], LR[34]. The first evaluation is performed on the

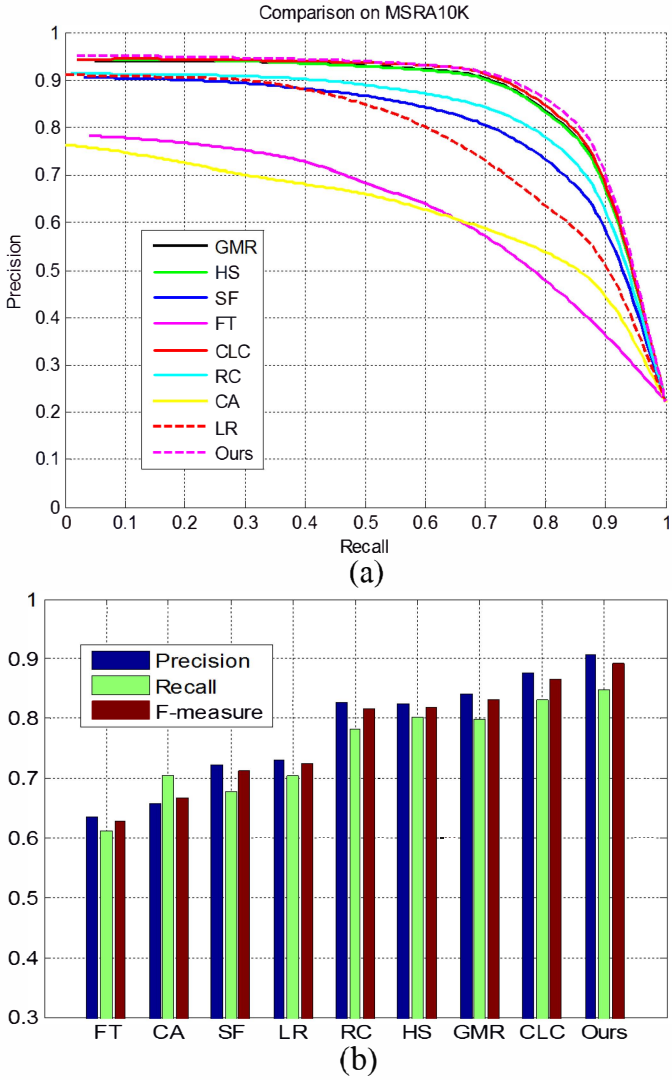

Fig. 3. Comparison of the proposed method with 8 state-of-the-art methods on the MSRA10K dataset. (a)Average precision-recall curves for different approaches. (b) Average precision, recall, and F-Measure using different approaches with an adaptive threshold.

MSRA10K dataset, and its results are depicted in Figure 3. Precision-recall curves are shown in Figure 3(a), and the average precision, recall, and F-measure comparison are shown in Figure 3(b). From figure 3(a), the proposed method performs as well as the CLC, HS, and GMR approaches. Our method has a slight superiority. Figure 3(b) shows that the proposed method has the highest average precision, recall, and F-measure among all the methods, which illustrate that our method excels the other 8 state-of-the-art approaches on MSRA10K dataset.

TABLE I
RUNNING TIME TEST RESULTS (SECONDS PER IMAGE).

Method	Ours	SF	GMR	HS	CLC	FT	RC	CA	LR
Time(s)	0.54	0.19	0.68	0.64	1.78	0.072	0.11	38	14
Code	M	M	M	EXE	M	C	C	M+C	M

We further evaluate the proposed algorithm used ECSSD dataset. The experiment process and evaluation metrics are similar to what we have done on basis of the MSRA10K dataset before. The precision-recall curves are shown in Figure 4(a), and the F-measure comparison is shown in Figure 4(b). Experimental results show that performance of different methods on MSRA10K is better than those on ECSSD dataset. Because the ECSSD dataset is more complex as it includes many images with multi-object and complex pattern. Figure 4(a) shows that the proposed algorithm has best precision-recall curve than all other methods and average precision, recall and F-measure also have obvious superiority than other methods which is shown in figure 4(b). Our method achieves the best performance on ECSSD dataset compared to other 8 state-of-the-art methods.

D. Running Time

We use 64-bit PC with Intel Core i5-4570 CPU @ 3.2 GHz and 8GB RAM to for running time test. The average running time is computed by using ASD dataset. We choose the methods that have close performance to the proposed algorithm, and the results are shown in Table 1. Where "C" stands for "C/C++", "M" stands for "Matlab", "M+C" stands for a mixture of "Matlab" and "C/C++", "EXE" means executable file. Our method still outperforms them considering the overall evaluation performances while it is slower than SF, FT, and RC.

V. CONCLUSION

This paper presents a novel bottom-up saliency method for detection, which takes advantage of both background prior and compactness prior. To achieve more accurate saliency results, the proposed approach optimizes the image boundary by removing erroneous boundary regions with a fixed threshold value. The saliency map obtained by ranking with background queries can be optimized with compactness prior. We first verify efficiency of the design options on ASD dataset, then we compare the proposed method with 8 state-of-the-art methods on MSRA10K and ECSSD datasets. The experiment results show that the proposed method outperforms 8 state-of-the-art saliency detection algorithms in terms of accuracy and robustness. As a future work, we will continue improving the performance of our algorithm, and apply it in more potential applications. Medical science and technology demands high accuracy for proper investigation and treatment of human beings. Therefore, our novel approach can find its potential applications in bioinformatics, medical sciences and medical image processing for tumor detection because of its high accuracy and robustness.

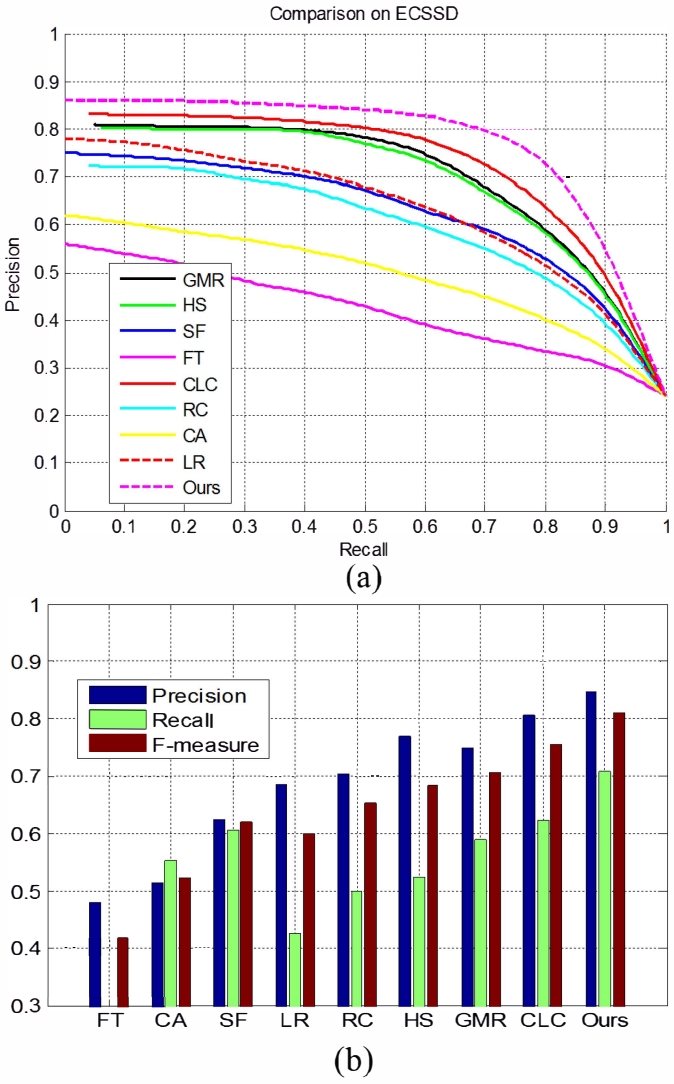

Fig. 4. Comparison of results, by the proposed method for this research with 8 state-of-the-art methods on the ECSSD dataset.(a) Average precision-recall curves for different approaches used. (b) Average precision, recall and F-Measure using different approaches with an adaptive threshold.

REFERENCES

[1] H. E. Pashler and S. Sutherland, *The psychology of attention*. MIT press Cambridge, MA, 1998, vol. 15.

[2] J. M. Henderson, "Human gaze control during real-world scene perception," *Trends in cognitive sciences*, vol. 7, no. 11, pp. 498–504, 2003.

[3] H. E. Egeth and S. Yantis, "Visual attention: Control, representation, and time course," *Annual review of psychology*, vol. 48, no. 1, pp. 269–297, 1997.

[4] J.-Y. Zhu, J. Wu, Y. Xu, E. Chang, and Z. Tu, "Unsupervised object class discovery via saliency-guided multiple class learning," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 4, pp. 862–875, 2015.

[5] B. Alexe, T. Deselaers, and V. Ferrari, "Measuring the objectness of image windows," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2189–2202, 2012.

[6] H. Fu, Z. Chi, and D. Feng, "Attention-driven image interpretation with application to image retrieval," *Pattern Recognition*, vol. 39, no. 9, pp. 1604–1621, 2006.

[7] X. Yang, X. Qian, and Y. Xue, "Scalable mobile image retrieval by exploring contextual saliency," *IEEE Transactions on Image Processing*, vol. 24, no. 6, pp. 1709–1721, 2015.

[8] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1304–1318, 2004.

[9] L. Zhang, Y. Sun, T. Luo, and M. M. Rahman, "Note: A manifold ranking based saliency detection method for camera," *Review of Scientific Instruments*, vol. 87, no. 9, pp. 309–320, 2016.

[10] L. Wang, J. Xue, N. Zheng, and G. Hua, "Automatic salient object extraction with contextual cue," in *2011 International Conference on Computer Vision*. IEEE, 2011, pp. 105–112.

[11] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on*. IEEE, 2009, pp. 1597–1604.

[12] M. Mancas, L. Couvreur, B. Gosselin, B. Macq *et al.*, "Computational attention for event detection," in *Proc. Fifth Intl Conf. Computer Vision Systems*, 2007.

[13] J. Shi, Q. Yan, L. Xu, and J. Jia, "Hierarchical image saliency detection on extended cssd," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 4, pp. 717–729, 2016.

[14] L. Itti, C. Koch, E. Niebur *et al.*, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[15] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in neural information processing systems*, 2006, pp. 545–552.

[16] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," *IEEE Transactions on Pattern analysis and machine intelligence*, vol. 33, no. 2, pp. 353–367, 2011.

[17] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proceedings of the 14th ACM international conference on Multimedia*. ACM, 2006, pp. 815–824.

[18] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2015.

[19] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1915–1926, 2012.

[20] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, and N. Crook, "Efficient salient region detection with soft image abstraction," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1529–1536.

[21] ——, "Efficient salient region detection with soft image abstraction," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1529–1536.

[22] Z. Ren, S. Gao, L.-T. Chia, and I. W.-H. Tsang, "Region-based saliency detection and its application in object recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 5, pp. 769–779, 2014.

[23] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8.

[24] C. Guo, Q. Ma, and L. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform," in *Computer vision and pattern recognition, 2008. cvpr 2008. ieee conference on*. IEEE, 2008, pp. 1–8.

[25] J. Li, M. D. Levine, X. An, X. Xu, and H. He, "Visual saliency based on scale-space analysis in the frequency domain," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 4, pp. 996–1010, 2013.

[26] J. Li, L.-Y. Duan, X. Chen, T. Huang, and Y. Tian, "Finding the secret of image saliency in the frequency domain," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 12, pp. 2428–2440, 2015.

[27] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature reviews neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.

[28] L. Zhang, L. Yang, and T. Luo, "Unified saliency detection model using color and texture features," *Plos One*, vol. 11, no. 2, 2016.

[29] J. Yang and M.-H. Yang, "Top-down visual saliency via joint crf and dictionary learning," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2296–2303.

[30] A. Borji, "Boosting bottom-up and top-down visual features for saliency estimation," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 438–445.

[31] ——, "Boosting bottom-up and top-down visual features for saliency estimation," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 438–445.

[32] G. Zhu, Q. Wang, and Y. Yuan, "Tag-saliency: Combining bottom-up and top-down information for saliency detection," *Computer Vision and Image Understanding*, vol. 118, pp. 40–49, 2014.

[33] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf, "Ranking on data manifolds," *Advances in neural information processing systems*, vol. 16, pp. 169–176, 2004.

[34] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 853–860.

[35] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3166–3173.

[36] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[37] C. Li, Y. Yuan, W. Cai, Y. Xia, and 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.

[38] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 733–740.

[39] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1155–1162.