

## Salient Region Detection by Integrating Intrinsic and Extrinsic Cues without Prior Information

Ji Ma<sup>1,2\*</sup>, Jingjiao Li<sup>1</sup>, Zhenni Li<sup>1</sup> and Jiao Jiao<sup>3</sup>

<sup>1</sup> School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China

<sup>2</sup> College of Information, Liaoning University, Shenyang 110036, China

<sup>3</sup> Metropolitan Technology Consulting Ltd, Orpington, BR6 9QR, United Kingdom

Received 10 January 2017; Accepted 29 June 2017

### Abstract

In recent years, independence on prior information and the efficient integration of multiple visual cues have become an important topic in salient region detection. A salient region detection framework that integrates intrinsic and extrinsic visual cues was proposed in this study to remove the limitations of prior information. First, salient prior map was generated using statistical texture representations and by classifying superpixels into three classes of background, salient regions, and unknown regions. Second, the final saliency map was computed by integrating multi-channel color features and color contrast saliency factors based on self-generating salient prior map. Finally, the experiment was conducted on popular benchmark datasets, namely, Microsoft Research Asia (MSRA) and Extended Complex Scene Saliency Dataset (ECCSD), using two standard criteria: precision-recall rate and F-measure rate. Results demonstrate that the proposed method is more competitive than methods based on prior information, such as dense and sparse reconstruction (DSR), absorbing Markov chain (MC), and robust background detection (RBD). The proposed method achieves an 8% reduction in computing complexity compared with DSR. Compared with the learning-based method without prior information, the performance of the proposed method resembles that of discriminative regional feature integration (DRFI) and its complexity is reduced by 70%. This study provides a novel method to improve the performance of salient region detection and to avoid prior information.

*Keywords:* Salient region detection, Random forest, Surperpixel

### 1. Introduction

Digital image is becoming an increasingly important resource in gaining information and knowledge as computer technology develops and mobile terminal becomes popular. The quick and accurate extraction of the key areas of an image and the preferential assignment of computing resources can effectively improve the efficiency and accuracy of image processing using salient region detection. Thus, the technology of salient region detection is widely used in image segmentation [1], object detection [2], and image retrieval [3].

Distinguishing foreground (salient region) from background is a key step in salient region detection, and finding the visual characteristics for distinguishing foreground from background is the key research content. Some intrinsic cues were proposed that are only extracted from the input image. However, Borji et al. argued that salient region and background may share some common visual attributes among images, and the intrinsic cues lack the capacity to distinguish these attributes [4]. Therefore, extrinsic cues, including manual annotations, depth map, or the similarities between images, were incorporated to

facilitate salient region detection in images.

Intrinsic or extrinsic cues must be represented by quantifiable representations, such as structure, contrast, texture, color, size or shape. Color attribute is most widely used as a single representation among these representations. The study [5] found that structure and texture representations can distinguish salient region from background. Methods that integrate multiple visual cues were proposed because single visual cue and single representation have their own defects. Perazzi et al. proposed a Gaussian filtering framework that integrates two intrinsic cues, namely, color representation and spatial distribution, as uniqueness and compactness, respectively [6]. However, color contrast and compactness representations cannot easily distinguish similar colors between salient region and background. Methods based on prior information were proposed to improve salient region detection. Remarkable results have been achieved by prior background methods using boundary and connectivity priors derived from probable backgrounds in natural images [7]. The prior information are primarily motivated by psychophysical observations, that is, salient objects have distance from image boundary and most background regions always connect to one another along the boundary. However, prior information limits the application range of salient region detection. Thus, the integration of multiple representations without prior information is the research direction of salient region detection.

\*E-mail address: horsemaji@126.com

ISSN: 1791-2377 © 2017 Eastern Macedonia and Thrace Institute of Technology. All rights reserved.

doi:10.25103/jestr.103.24

This study proposes a multi-stage framework of salient region detection. Different visual cues are used in different stages to utilize the advantages of different features. In the gradual pattern of salient region detection, the salient prior map computed with classification algorithm is used instead of the priori model of artificial design.

## 2. State-of-the-art

Distinguishing and segmenting salient regions from background are key steps when detecting salient regions. Searching for visual characteristics, which have discriminative abilities of salient region detection, is key in distinguishing salient regions from background [4] - [5]. Color contrast and spatial distribution representations extracted from the input image itself as the intrinsic cues are applied in salient region detection. Although salient regions and background may share some common visual representations, intrinsic cues are not good at distinguishing these representations. Therefore, the study of extrinsic cues with user annotations or statistical information of similar images received increased attention.

Methods about global regional color contrast used as intrinsic cue were widely studied in existing approaches [8] - [9]. Jiang et al. obtained effective result by modeling the salient characteristic in color space with an absorbing Markov chain [10]. In addition to color representation, other representations, such as texture [11] and structure [12], were also proposed for salient region detection. These methods use single visual feature simply and lack the integration of multiple visual features. Methods that integrate multiple features were proposed because single representation has limits in distinguishing salient regions. Features that include

global color spatial distribution, local multi-scale contrast, and regional center surround histogram distance were integrated in the method [13]. Cheng et al. evaluated the global contrast differences and spatial coherence using the Gaussian mixture model (GMM) to improve the accuracy of salient region detection [14]. Regional features with color and shape were considered in the approach [15]. These methods only focus on the integration of intrinsic visual cues and ignore the correlation of visual characteristics between images. Although machine learning methods were widely applied in other areas of computer vision, such as image classification, research in salient region detection are few. Learning-based algorithms for salient region detection were also proposed because labeled datasets for saliency detection were proposed recently. A supervised learning method [16] was performed using regional descriptors to build a saliency map with multilevel image segmentation. However, this method lacks intrinsic and extrinsic visual cue integration.

Prior information was defined by analyzing the human visual characteristics for salient region detection [4]. The background measure method was proposed as prior information that considers image boundary as the background and defines boundary connectivity to estimate background regions [9]. Li et al. used a set of background templates as prior information to construct the dense and sparse representation errors of each image region; they also computed the saliency map by integrating reconstruction errors [17]. In [10], salient regions were detected by integrating different representations with prior information known as center prior. Prior-based methods achieved outstanding performance on popular benchmark datasets. However, prior information limits the application fields of salient region detection.

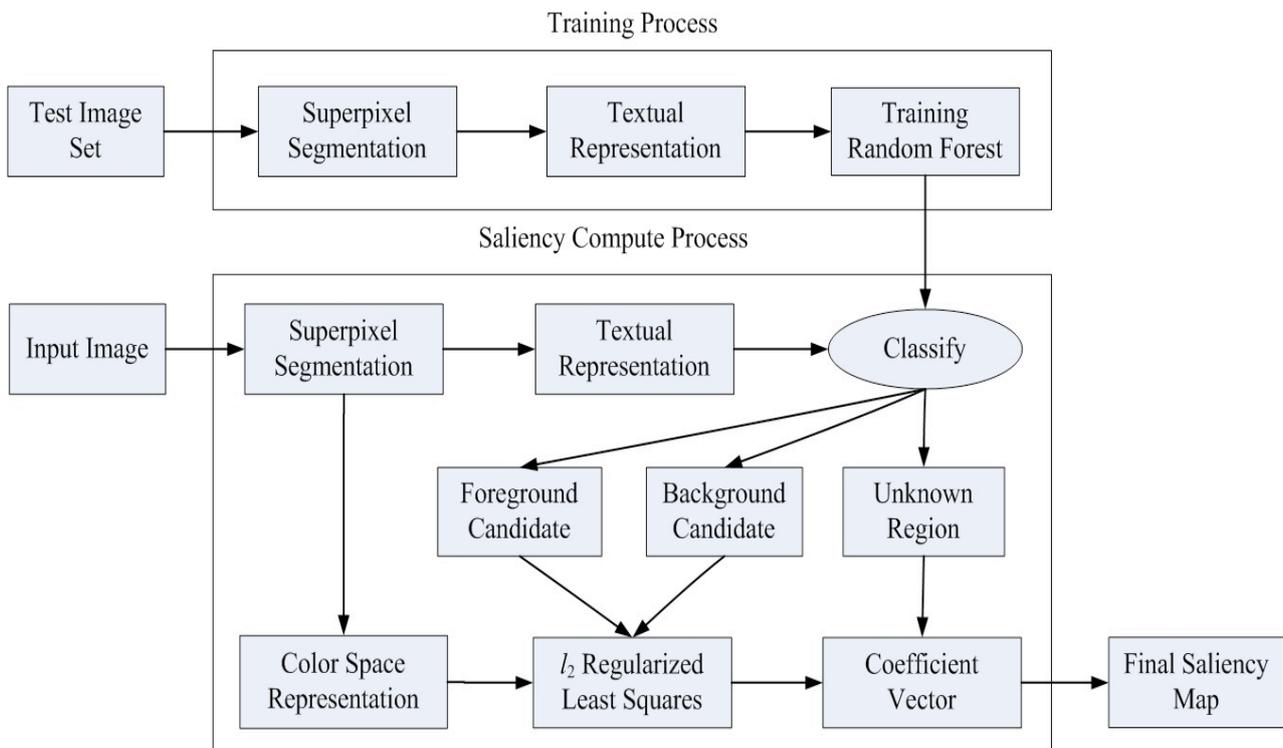


Fig. 1. Detailed approach of the proposed method

These studies mostly depend on prior information and integrate multiple features simply as a result of the lack of intrinsic and extrinsic cue fusions. This study considers the full advantage of intrinsic and extrinsic cues with multiple features, and implements the avoidance of prior information into a unified framework based on the perspective of integrating cues and features. This study uses the algorithm presented in [6] to segment an input image into regions; a regional saliency map is then computed in the superpixel level instead of pixel-wise level to reduce computational time. The approaches of J. Kim et al. [18] and C. Scharfenberger et al. [11] are closely related to our method. The method in [18], which integrates intrinsic and extrinsic cues, doesn't use statistic textural characteristic. The method in [11], which uses statistic textural characteristic, neglects the learning-based cue. The present study extends existing

methods and integrates intrinsic and extrinsic cues in multi-stage framework. Textural characteristic is used as an extrinsic cue to compute the salient prior map to remove dependence on prior information. The final saliency map is calculated by integrating the multi-channel color characteristics and color contrast feature using the salient prior map as prior information. This approach is robust in applications without prior information.

The rest of this study is organized as follows. Section 3 describes the statistic textural model used as an extrinsic cue for the salient prior map. A method is proposed, which integrates intrinsic cues to generate the final saliency map based on self-generating salient prior map. Section 4 presents the experimental results and discusses the performance with several state-of-the-art saliency methods. Section 5 summarizes and provides conclusions.

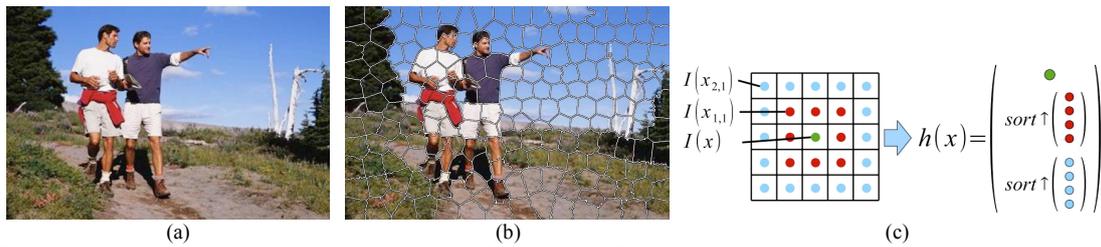


Fig. 2. Superpixel segmentation and textural representation. (a) Raw image (b) Superpixel segmentation (c) Textural representation

### 3. Methodology

#### 3.1 Generation of Salient Prior Map

The generation of salient prior map (SPM) from an image is described in detail in this section. This approach processes an image in the region level with superpixel segmentation and uses statistic textural representation with random forest regressor to classify superpixels. Similar to [18], the SPM consists of background candidate, foreground candidate, and unknown regions. The accuracy of salient region classification in SPM is further improved by benefiting from the class of some ambiguous unknown regions. Fig. 1 describes the detail approach of the proposed method.

##### 3.1.1 Superpixel Segmentation and Textual Representation

The features extracted from superpixels are effective and efficient for salient region detection according to [6]. Superpixels preserve the local image structure, such as region contours between superpixels. Figs. 2a and 2b show that the input image  $I$  is segmented to form superpixels. Simple linear iterative cluster (SLIC) [19] is used because of its high performance and low computational cost.

A texture representation model represents the textural characteristics of local superpixels, such as that in [11]. Let  $I$  be the  $M \times N$  image, and  $I(x)$  is the intensity of a pixel  $x$  in image  $I$ . Given an adjacent area  $\Omega$  with pixel  $x$  in the center, the local textural representation  $h_c(x)$  for each color channel  $c$  is defined as:

$$h_c(x) = I_c(x) \text{sort}_{\uparrow} \{I_c(x_{1,j})\} \text{sort}_{\uparrow} \{I_c(x_{2,j})\} \dots \text{sort}_{\uparrow} \{I_c(x_{n,j})\} \quad (1)$$

where  $x_{i,j}$  is the  $j^{\text{th}}$  pixel position in the  $i^{\text{th}}$  pixel radial layer around  $x$ , and  $\text{sort}_{\uparrow}$  denotes sorting in the intensity by ascending order. Fig. 2c shows the process for single color channel. Similar to [11], Principal Component Analysis (PCA) is used to reduce the dimension of  $h(x)$  with the  $u$  principal components:

$$w_x = \{\Psi_i(h(x)) | 1 \leq i \leq u\} \quad (2)$$

where  $\Psi_i$  is the  $i^{\text{th}}$  principal component of  $h(x)$ . In this study,  $u$  is selected to represent 95% of the variance of all textural representations. Then, the set of local texture representations with the image  $I^{M \times N}$  is obtained:

$$W = \{w_1, w_2, w_3, \dots, w_x, \dots, w_{M \times N}\} \quad (3)$$

where  $w_i$  is the textural representation of the  $i^{\text{th}}$  pixel in image  $I^{M \times N}$ . For the textural representations of the aforementioned superpixels, a set  $W^e$  of  $J$  superpixels is defined as:

$$W^e = \{w_{e1}, w_{e2}, w_{e3}, \dots, w_{eJ}\} \quad (4)$$

where  $w_{e\nu}$ ,  $1 \leq \nu \leq J$  is the texture feature of superpixel  $e\nu$ .  $w_{e\nu}$  is computed by averaging all  $U$  local textural representations  $w_x$  within a superpixel  $e\nu$ .

$$w_{e\nu} = \frac{1}{U} \sum_x w_x, x \in e\nu \quad (5)$$

### 3.1.2 Definition of Saliency with Texture Representation

Salient regions with highly unique and distinctive textural characteristics are considered based on the consensus of saliency. A quantitative metric is used to distinguish texture patterns relative to one another. The model of statistical textural distinctiveness demonstrated good performance, wherein a region is salient when the textural pattern commonality is low. Then, the full utilization of the statistical textural distinctiveness is described to discern underlying saliency.

Let  $w_{ei}$  and  $w_{ej}$  denote a pair of texture elements, where  $w_{ej}$  is a noisy observation of another  $w_{ei}$ , and  $w_{ei}$  and  $w_{ej}$  differ only in an additive random component  $\eta_{i,j}$ :

$$w_{ej} = w_{ei} + \eta_{i,j} \quad (6)$$

$\eta_{i,j}$  is a random field with probability distribution  $P(\eta_{i,j})$  used to show the differences between  $w_{ei}$  and  $w_{ej}$  with the inherent texture variability.  $\eta_{i,j}$  is used as the Gaussian random field with zero-mean and variance  $\delta^2$  corresponds to the variance of the  $L_p$ -norm ( $\delta^2 = var(\|w_{ei} - w_{ej}\|_p)$ ) between  $w_{ei}$  and  $w_{ej}$ . The probability of  $w_{ei}$  is then defined as a noisy observation of  $w_{ej}$  as:

$$P(w_{ei} | w_{ej}) = \prod_k P(w_{ei,k} | w_{ej,k}) \quad (7)$$

where  $P(w_{ei,k} | w_{ej,k})$  is the probability of element  $k$  in  $w_{ei}$  being a noisy observation of the corresponding element  $k$  in  $w_{ej}$ . Let  $\beta_{i,j}$  replace  $P(w_{ei} | w_{ej})$ , a textural distinctiveness matrix  $T^{n \times n}$  is constructed based on the  $n$  superpixels. Each entry  $t_{i,j}$  in the matrix is associated with the statistical textural distinctiveness  $\beta_{i,j}$  between a pair of textural representations  $w_{ei}$  and  $w_{ej}$ .

### 3.1.3 Estimation of SPM

The SPM of an image is estimated using the aforementioned textural distinctiveness matrix  $T^{n \times n}$  based on the following assumption: statistical textural distinctiveness and probability of occurrence between regions. The probability of occurrence is defined as the number of pixels in the corresponding region. A classification algorithm is used to estimate the SPM in contrast to the threshold-based method [11]. In this study, the random forest classification [20] is used to classify superpixels into foreground, background, and unknown regions because of its efficiency in large databases and its generalization ability. Vector  $\alpha_i$ , which represents the salient feature for the region  $S_i$ , is defined as follows:

$$\alpha_i = (|S_1| \cdot \beta_{i,1}, |S_2| \cdot \beta_{i,2}, \dots, |S_j| \cdot \beta_{i,j}, \dots, |S_n| \cdot \beta_{i,n}) \quad (8)$$

where  $|S_j|$  is the number of pixels in  $S_j$ . Images with ground truth labels from the Microsoft Research Asia (MSRA) dataset [15] are used. The random forest classification provided by Becker et al. [21] is used. The relatively reliable superpixels of salient and non-salient superpixels are considered as foreground or background, respectively, and the ambiguous superpixels are considered unknown. The classification of ambiguous regions can make the location of salient regions more reliable than binary saliency maps without unknown regions. The threshold values  $T_{fore} = 1$  and  $T_{back} = -1$  are used to decide which class each superpixel belongs to through the response value extracted from the classifier. If a superpixel obtains a response value over  $T_{fore}$ , then the superpixel belongs to the foreground. However, if the value is lower than  $T_{back}$ , then the superpixel belongs to the background; otherwise, it is classified unknown.

## 3.2 Final Saliency Estimation from SPM

The next stage uses color representations to estimate the final saliency map based on SPM.

### 3.2.1 Color Representation for Saliency

Colors are important visual attributes in the visual system of a human. In this section, color representations are used to generate the final saliency map. Similar to [18], many different color spaces, including YUV (The Y denotes the luminance signal, and U, V denote the two chrominance signals), hue saturation value (HSV), and commission internationale de l'eclairage  $L^*, a^*, b^*$  (CIE Lab), are integrated. Searching for a linear combination of color coefficients in multiple color spaces may be a good choice because searching for a suitable color space that strongly correlates to human perception is difficult. The values in each channel of color spaces between [0, 1] are normalized to construct a high-dimensional matrix:

$$T = \begin{bmatrix} R_1^{\gamma_1} & R_1^{\gamma_2} & R_1^{\gamma_3} & G_1^{\gamma_1} & \dots \\ R_2^{\gamma_1} & R_2^{\gamma_2} & R_2^{\gamma_3} & G_2^{\gamma_1} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ R_N^{\gamma_1} & R_N^{\gamma_2} & R_N^{\gamma_3} & G_N^{\gamma_1} & \dots \end{bmatrix} \in R^{N \times l} \quad (9)$$

where  $R_i$  and  $G_i$  are mean pixel value of the  $i^{th}$  superpixel for corresponding color channel in the image. By utilizing the salient region and background color samples in the SPM, estimating an optimal linear combination of color coefficients is formulated as a  $l_2$  regularized least squares problem that minimizes:

$$\min_{\theta} \|(Q - \tilde{T}\theta)\|_b^2 + \lambda \|\theta\|_b^2 \quad (10)$$

where  $\theta \in R^l$  is the coefficient vector to be estimated,  $\lambda$  is a weighting parameter to control the magnitude of  $\theta$ , and  $\tilde{T}$  is a  $M \times l$  matrix. Each row of  $\tilde{T}$  corresponds to color representations in the foreground or background candidate regions:

$$\tilde{T} = \begin{bmatrix} R_{F_1}^{\gamma_1} & R_{F_1}^{\gamma_2} & R_{F_1}^{\gamma_3} & G_{F_1}^{\gamma_1} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ R_{F_x}^{\gamma_1} & R_{F_x}^{\gamma_2} & R_{F_x}^{\gamma_3} & G_{F_x}^{\gamma_1} & \dots \\ R_{B_1}^{\gamma_1} & R_{B_1}^{\gamma_2} & R_{B_1}^{\gamma_3} & G_{B_1}^{\gamma_1} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ R_{B_y}^{\gamma_1} & R_{B_y}^{\gamma_2} & R_{B_y}^{\gamma_3} & G_{B_y}^{\gamma_1} & \dots \end{bmatrix} \quad (11)$$

where  $F_i$  is the  $i^{th}$  foreground candidate superpixel and  $B_j$  is the  $j^{th}$  background superpixel that are classified at SPM generation step, respectively.  $M$  is the number of superpixels in the foreground/background candidate regions and  $N$  is the number of total superpixels in the image ( $M \ll N$ ). Let  $x$  and  $y$  denote the number of superpixels in foreground and background, respectively, such that  $M = x + y$ .  $Q$  is an  $M$  dimensional vector with a value equal to 1 and 0 when a superpixel belongs to the foreground and background candidate, respectively:

$$Q = \left[ \underbrace{1 \ 1 \ \dots \ 1}_x \ \underbrace{0 \ 0 \ \dots \ 0}_y \right]^T \in R^{M \times 1} \quad (12)$$

The  $l_2$  regularized least square problem can be readily minimized with respect to  $\theta$  as  $\theta^*$  because the number of superpixels is greater than the dimensions of the coefficient vector. In all experiments,  $\lambda = 0.05$  is used to attain the best result.

$$\theta^* = (\tilde{T}^T \tilde{T} + \lambda I)^{-1} \tilde{T}^T Q \quad (13)$$

### 3.2.2 Computation of Final Saliency Map

The saliency map that can be constructed after  $\theta^*$  is obtained:

$$\phi(w_i) = \sum_{j=1}^I T_{ij} \theta_j^*, \quad i = 1, 2, \dots, N \quad (14)$$

With the  $l_2$  regularizer, the constructed saliency map is more reliable for the foreground and background superpixels in the SPM. However, the  $l_2$  regularized least square algorithm is sensitive to noise. Thus, the result must be adjusted with another cue. Motivated by the method [4], the salient contrast factor  $U_i$  and background contrast factor  $S_i$  are defined as:

$$U_i = \sum_{j=1}^N \|c_i - c_j\|^2 \quad (15)$$

and

$$S_i = \sum_{k=1}^M \|c_i - c_k\|^2 \quad (16)$$

where  $N$  and  $M$  are the numbers of salient and background superpixels in the SPM, respectively. Salient contrast factor  $U_i$  is generally defined as the contrast of the  $i^{th}$  superpixel

given its color in CIELab  $c_i$  compared with all other salient superpixels in the SPM. Background contrast factor  $S_i$  is defined in a similar way. The small value of  $U_i$  implies that the  $i^{th}$  superpixel more likely belongs to salient region. Similarly, the small value of  $S_i$  implies that the  $i^{th}$  superpixel more likely belongs to the background.  $U_i$  and  $S_i$  are normalized to adjust Eq. (14) by:

$$\phi(w_i) = \begin{cases} \alpha \sum_{j=1}^I T_{ij} \theta_j^* & , \left| 1 - \frac{U_i}{S_i} \right| \leq \delta \\ \max(1 - U_i, S_i) \sum_{j=1}^I T_{ij} \theta_j^* & , \left| 1 - \frac{U_i}{S_i} \right| > \delta \end{cases}, \quad i = 1, 2, \dots, N \quad (17)$$

where  $\delta$  is a small value that represents the difference between  $U_i$  and  $S_i$ . When  $U_i$  is similar to  $S_i$ , then salient contrast factor and background contrast factor cannot help distinguish the salient region from background. Thus,  $\alpha = 0.5$  is set as a balance factor.

## 4 Result Analysis and Discussion

The performances of the proposed method were evaluated and compared against previous algorithms, such as the methods proposed by Jiang et al., which is the discriminative regional feature integration (DRFI) [16], Li et al., which is the dense and sparse reconstruction (DSR) [17], Jiang et al., which is the absorbing Markov chain (MC) [10], J. Kim et al., which is the high-dimensional color transform (HDCT) [18], and Zhu et al., which is the robust background detection (RBD) [9], on popular benchmark datasets, such as MSRA [15] and Extended Complex Scene Saliency Dataset (ECCSD) [22]. In the present study, precision–recall rate and F-measure rate, which were described by criteria [7], were used to evaluate our salient region detection algorithm.

The threshold was adjusted from 0 to 255 to generate saliency map. Moreover, the precision–recall curve was plotted with the x-axis and the y-axis as the recall and precision rates, respectively. The precision–recall curve of some state-of-the-art algorithms, including our algorithm, is shown in Fig. 3. The results show that our method is more competitive than previous methods.

F-measure rate coordinates Precision and Recall rates for comprehensive evaluation are defined below:

$$F_\beta = \frac{(1 + \beta^2) Precision \times Recall}{\beta^2 Precision + Recall} \quad (18)$$

Similar to [22], the reconciled parameter  $\beta$  was set as  $\beta^2 = 0.3$ . The F-measure curve was plotted with the x-axis as the threshold and the y-axis as the F-measure rate. Fig. 4 shows that the F-measure curves of the state-of-the-art algorithms show that our approach has a competitive global performance compared with other methods.

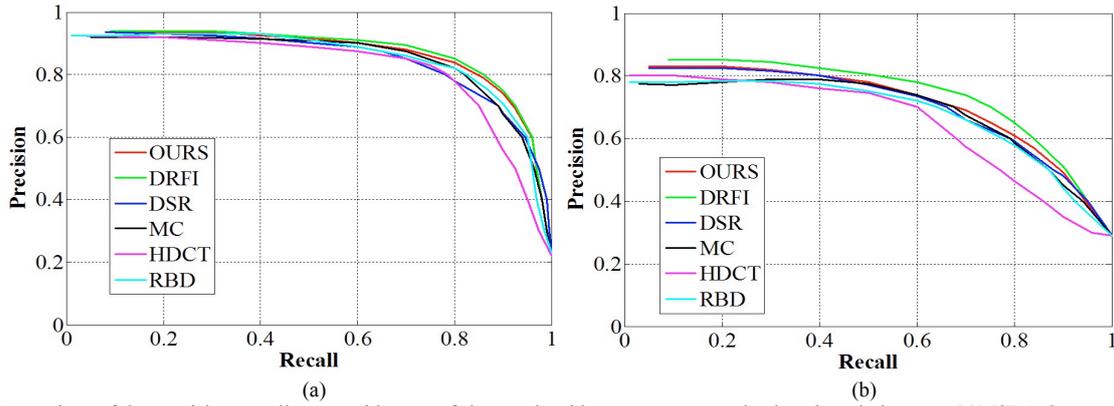


Fig. 3. Comparison of the precision–recall curve with state-of-the-art algorithms on representative benchmark datasets: (a)MSRA dataset (b)ECSSD dataset

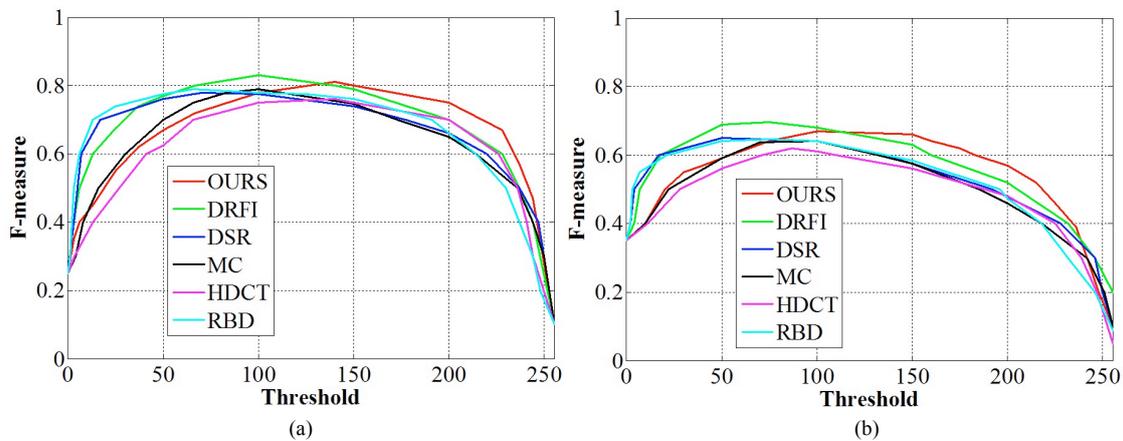


Fig. 4. Comparison of the F-measure curve with five state-of-the-art algorithms on representative benchmark datasets: (a)MSRA dataset (b)ECSSD dataset

The average computational times for the compared methods are shown in Table 1 to further demonstrate the performance of our method. The running environment is a laptop with an Intel Dual Core i5-2500K 3.30 GHz CPU, 8.00GB RAM and the software environment is MATLAB2012b. Results show that our approach achieved competitive performance compared with other methods. The superpixel generation step (about 0.9 s) and feature vector generation step (about 4.1 s) for SPM were the most time-consuming steps based on the computational complexity of our proposed method. DRFI[16] and HDCT[18] are state-of-

the-art methods without prior information. The performance of our method is inferior to DRFI but better than HDCT. However, computation complexity was reduced by 70% compared with DRFI. This algorithm compromises balances in accuracy and complexity. The performance of our method is comparable to methods based on prior information, such as DSR[17], MC[10] and RBD[9]. An 8% reduction in computing complexity for DSR is observed in our method. Although prior information produces excellent results in benchmark datasets, they have limitations in real applications.

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

Method	Our method	DRFI[16]	DSR[17]	MC[10]	HDCT[18]	RBD[9]
Time(s)	5.6	19.3	6.1	0.1	3.3	0.3

### 5. Conclusions

A novel multi-stage method was presented in this study to address the problem of prior information dependency and lack of fusing intrinsic and extrinsic cues in salient region detection. This method calculated the SPM and the final saliency map and corrected noise error by integrating color features and statistical texture features based on learning and by using color contrast features, respectively. Finally, experiments were performed on two popular benchmark datasets. The following conclusions are drawn:

(1) Statistical texture features are rotational-invariant and effective for the discrimination of salient regions.

Complexity is decreased effectively with superpixel segmentation because pixel-level statistical texture features are relatively complex.

(2) When classification is used instead of the fixed threshold, the accuracy of SPM is effectively improved. Reliable reference is provided for subsequent calculation of saliency map.

(3) The multi-stage integration of intrinsic and extrinsic cues effectively takes full advantage of different cues. The experiments prove that this method is very effective in improving the precision of salient region detection.

The algorithm presented in this study eliminates the dependence on prior information in salient region detection,

improves the adaptability of the algorithm in real application, and provides good preprocessing performance for other computer vision applications. Future work should examine parallel algorithms to improve the efficiency of salient region detection. Consequently, heuristic optimization

algorithm should be considered to solve the noise-sensitive problem of least squares optimization better.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence



## References

1. Z. Liu, R. Shi, L. Shen, Y. Xue, K. N. Ngan, and Z. Zhang, "Unsupervised salient object segmentation based on kernel density estimation and two-phase graph cut". *IEEE Transactions on Multimedia*, 14(4),2012,pp.1275-1289.
2. Z. Li and L. Itti, "Saliency and gist features for target detection in satellite images". *IEEE Transactions on Image Process*, 20(7),2011,pp. 2017-2029.
3. Boato, G., Dang-Nguyen, DT., Muratov, O. et al, "Exploiting visual saliency for increasing diversity of image retrieval results". *Multimed Tools Appl*,75(10),2016 , pp. 5581-5602.
4. A. Borji, M.-M. Cheng, H. Jiang, and J. Li, "Salient object detection: A survey". *Eprint Arxiv*,16(7),2014,pp.3118-3144.
5. A. Borji, M.-M. Cheng, H. Jiang, and J. Li. (2015). "Salient object detection: A benchmark". *IEEE Transactions on Image Processing*, 24(12) , 2015,pp.5706-5722.
6. F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection". In: *2012 Conference on Computer Vision and Pattern Recognition*, Rhode island, USA: IEEE,2012,pp. 733-740.
7. Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," In: *2012 European Conference on Computer Vision*, Florence, Italy:Springer,2012,pp. 29-42.
8. M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection". *IEEE Transactions on Pattern Analysis and Machine Intelligence* ,37(3),2015,pp.569-582.
9. W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection". In: *2014 Conference on Computer Vision and Pattern Recognition*, Columbus, USA: IEEE, 2014,pp. 2814-2821.
10. B. Jiang, J. Zhang, H. Lu, C. Yang, and M.-H. Yang, "Saliency detection via absorbing Markov Chain". In: *2013 International Conference on Computer Vision*, Sydney, Australia:IEEE,2013,pp. 1665-1672.
11. C. Scharfenberger, A. Wong, K. Fergani, J. S. Zelek, and D. A. Clausi, "Statistical textural distinctiveness for salient region detection in natural images". In: *2013 Conference on Computer Vision and Pattern Recognition*, Portland, USA: IEEE,2013,pp. 979-986.
12. K. Shi, K. Wang, J. Lu, and L. Lin, "Pisa: Pixelwise image saliency by aggregating complementary appearance contrast measures with spatial priors". In: *2013 Conference on Computer Vision and Pattern Recognition*, Portland, USA: IEEE,2013, pp. 2115-2122.
13. T. Liu, J. Sun, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object". In: *2007 Conference on Computer Vision and Pattern Recognition*, Minneapolis, USA: IEEE,2007, pp. 1-8.
14. M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, and N. Crook, "Efficient salient region detection with soft image abstraction". In: *2013 International Conference on Computer Vision*, Sydney, Australia: IEEE,2013, pp. 1529-1536.
15. P. Mehrani and O. Veksler, "Saliency segmentation based on learning and graph cut refinement". In: *2010 British Machine Vision Conference*, Aberystwyth, UK:Springer,2010, pp. 1-12.
16. H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach". In: *2013 Conference on Computer Vision and Pattern Recognition*, Portland, USA: IEEE,2013, pp. 2083-2090.
17. X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction". In: *2013 International Conference on Computer Vision*, Sydney, Australia:IEEE,2013, pp. 2976-2983.
18. J. Kim, D. Han, Y.-W. Tai, and J. Kim, "Salient region detection via high-dimensional color transform". In: *2014 Conference on Computer Vision and Pattern Recognition*, Columbus, USA: IEEE, 2014,pp. 883-890 .
19. 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* , 34(11),2012, pp. 2274-2282.
20. L. Breiman, "Random Forests". *Machine Learning*,45(1),2001,pp 5-32.
21. C. Becker, R. Rigamonti, V. Lepetit, and P. Fua, "Supervised feature learning for curvilinear structure segmentation". In: *2013 Conference on Medical Image Computing and Computer Assisted Intervention*, Nagoya, Japan: Springer,2013,pp. 526-533.
22. Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection". In: *2013 Conference on Computer Vision and Pattern Recognition*, Portland, USA: IEEE,2013, pp. 1155-1162.