

Exploiting contrast cues for salient region detection

Jie Niu^{1,2} · Xiongzhu Bu¹ · Kun Qian³

Received: 13 September 2015 / Revised: 31 January 2016 / Accepted: 3 March 2016 /
Published online: 15 March 2016
© Springer Science+Business Media New York 2016

Abstract Visual saliency detection is an important cue used in human visual system, which can offer efficient solutions for both biological and artificial vision systems. Although there are many saliency detection models that can achieve good results on public datasets, the accuracy and reliability of salient object detection models still remains a challenge. For this reason, a novel effective salient region detection model is presented in this paper. Based on the principle that a combination of global statistics and surrounding contrast saliency operators can yield even better results than just using either alone, we use a histogram-based contrast method to calculate the global saliency values in an opponent color space. At the same time, we partition the input image into a set of regions, and the regional saliency is detected by considering the color isolation with spatial information and textural distinctness simultaneously. The final saliency is obtained based on a weighted fusion of the two saliency results. The experimental results from three widely used databases validate the efficacy of the proposed method in comparison with fourteen state-of-the-art existing methods.

Keywords Visual attention · Salient region detection · Contrast measure · Saliency map

1 Introduction

Visual saliency is a set of mechanisms that help to optimize the search processes inherent in moment-to-moment perception and cognition. Since it can locate the more interesting regions in a scene and reduce the computational load, modeling visual saliency can offer efficient

✉ Xiongzhu Bu
213101000013@njust.edu.cn

Jie Niu
njrogel@gmail.com

¹ School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing, China

² School of Electrical and Electronic Engineering, Changzhou College of Information Technology, Changzhou, China

³ School of Automation, Southeast University, Nanjing, China

solutions for biological and artificial vision systems, such as image segmentation [6, 15, 19], image classification [26], image retrieval [29] and video compression [10].

Saliency region detection methods can be divided into two categories: bottom-up stimulus-driven [11, 14] and top-down task-driven approaches [17, 31]. Bottom-up saliency is the process of identifying salient points or detecting salient regions that typically attract people's attention. In contrast, in top-down, the object that is being searched for is known, such as a template-based search. Most existing saliency detection methods are based on a bottom-up computational framework. These models fall into two general categories: 1) models that attempt to predict human fixations (e.g., IT [14], AIM [5], SUN [36]) and 2) models that aim to identify the rarity of either global or local contrasting features over the entire scene (e.g., Zhang et al. [37], SR [12], and Liu et al. [20]). Specifically, Itti et al. [14] combined multi-scale image features, computed using a set of center-surrounding operations, into a single topographical saliency map. Bruce et al. [5] proposed a model built entirely on computational constraints, which has achieved good success in predicting fixation patterns. Zhang et al. [36] used a Bayesian framework to calculate visual saliency based on natural image statistics. All the above methods focus on individual pixels, and the saliency map is always blurred. However, the true usefulness of a saliency map is determined based on its application, and people tend to pay more attention to regions within an entire scene, rather than independent pixels. Saliency region/object detection theory has been revived due to the development of a number of computer vision and computer graphics applications. Jiang et al. [16] ingeniously presented a saliency estimation algorithm called the discriminative regional feature integration (DRFI), which regards saliency estimation as a regression problem, and their algorithm achieves very good results in comparison with other methods. Zhang et al. [37] utilized the Markov chain method to get the saliency map, and their approach can consistently locate multiple objects. However, although their algorithm is based on super-pixels, it is still time-consuming due to its intrinsic characteristics. Hou and Zhang [12] proposed a novel method to build corresponding full resolution saliency maps in the spatial domain. Liu et al. [20] built a conditional random field to effectively combine multiple features for salient object detection and obtained good results.

Based upon the above analysis, it can be observed that most models attempt to find certain areas that stand out of the scene, whether in the frequency domain or the time domain. We also know that the human visual system is particularly sensitive to contrast (such as color, orientation, and pattern) [25]. From this perspective, the problem can be considered from two angles: global statistics and surrounding contrast. Global statistics contrast methods evaluate the saliency of an image in the frequency domain, or investigate the statistical characteristics of more unique and important features in the time domain. Achanta et al. [2] exploited contrast information relating to color and luminance to propose a frequency-tuned method that defines the pixel saliency of the entire image. Li et al. [18] validated a frequency domain-based saliency detector based on a scale-space analysis. Luo et al. [21] used a PCA-based method to extract pre-defined features to obtain the global salient information of the salient object. Cheng et al. [6] utilized 3D color space to evaluate color contrast, and used histogram-based speed-up method to refine the saliency model, yielding good results. Since most global statistics contrast methods ignore the existing spatial relationship in the image, these methods have difficulty in distinguishing between similar colors, regardless of whether they belong to the foreground or the background. In contrast to global statistics methods, surrounding contrast methods incorporate spatial relationships into the regional-level contrast computation, and the saliency contrast computation usually assumes that areas close to the

current location play more important roles than areas that are further away [33]. Ma and Zhang [22] obtained a saliency model based on contrast analysis, and then extended the model using a fuzzy growth approach which achieved better results. Jiang et al. [15] used the difference between the color histogram of a region and its immediate neighboring regions to calculate the saliency score. Goferman et al. [9] used four principles to highlight salient objects along with their contextual information. Achanta et al. [1] proposed a simple and fast contrast based method to generate saliency maps, which used low-level features of luminance and color. Rahtu et al. [24] proposed a salient object segmentation method which is based on a statistical framework and local contrast of illumination, color, and motion information features. These methods are often very intuitive, and tend to produce higher values near the edges, but fail to uniformly highlight the entire salient region. Despite the achievements of global statistics and surrounding contrast methods, neither of these methods alone can achieve optimal performance. Better performance can be obtained by combining the two approaches and taking the best practices from each method.

In this paper, we propose a salient region detection method, which exploits the strength of both saliency operations. The first operation, global statistics, considers the color statistics information in an opponent color space (I, RG, BY). The second operation, surrounding contrast, considers both spatial and contrast information to evaluate the saliency of a patch with respect to all other patches in the image.

The remainder of this paper is organized as follows. The proposed approach is introduced in detail in Section 2. Experimental results and comparisons are presented in Section 3. Finally, conclusions are drawn in Section 4.

2 Proposed saliency model

In this work, the practice and advantage of using both global statistics and surrounding contrast saliency is reconsidered. As illustrated in Fig. 1, a smoothing operation is firstly performed to generate more homogeneous regions, and a histogram-based algorithm is also used to reduce the number of colors. The global statistics saliency is then computed in an opponent color space (I, RG, BY). On the other hand, a simple linear iterative clustering algorithm (SLIC) [3] is used to generate uniform superpixels, and the surrounding contrast saliency is then considered on two fronts: 1) Color contrast and spatial information. 2) Textural distinctness. Finally, the global statistics and surrounding contrast saliency are combined to obtain the overall saliency map.

2.1 Global statistics saliency generation

Global statistics-based methods depict the global contrast features of an entire image without losing local information, so they can assign an approximate saliency value to similar features, and uniformly highlight salient regions.

In this paper, we define a pixel-level saliency computational method for the input image. Specifically, the statistical characteristics of the histogram are used to incorporate the color coherence when calculating the saliency value of pixel I_c in image I as follows:

$$S(I_c) = \sum_{j=1}^N f_c D(I_c, I_j) \quad (1)$$

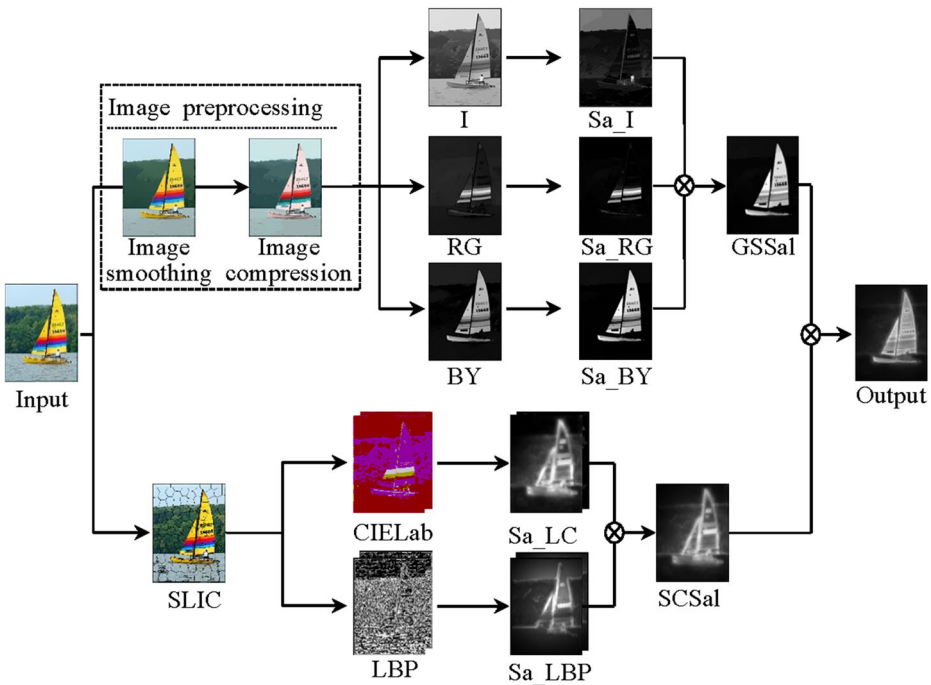


Fig. 1 Illustration of the main phases of our algorithm

where f_c is the frequency of color c in the image, $D(I_c, I_j)$ denotes the color distance metric between pixels I_c and I_j , and N is the number of distinct pixel colors.

2.1.1 Image preprocessing

Global statistics-based methods use the rarity of color information to directly define pixel saliency. From this perspective, pixels of the same color in an image will have the same saliency value. In order to finally obtain a uniform saliency map, the smoothing image of the input image is firstly found using gradient minimization [30], which results in a more homogeneous background. As discussed in [30], the smoothing parameter λ must be manually assigned a value and thus, this parameter may not be able to reach its optimal value. In this paper, we propose an automatic method to calculate λ based on the image entropy evaluation method. According to Shannon’s information theory, the entropy concept from thermodynamics can be used to quantify information. Therefore, we use Eq. (2) to measure the information capacity of an image:

$$H_{R_i} = -\sum_{v=0}^{255} p_v \ln p_v \tag{2}$$

where p_v is the probability of pixel intensity v within the image, estimated using a histogram. The entropy value can then be mapped to $[0, 9]$ based on the actual situation. In order to prevent the phenomenon of excessive smoothing from occurring, we use $\lambda = 0.005 * H$ to calculate the smoothing parameter.

Theoretically, in RGB color space, each color is usually chosen from a palette of 256^3 colors. Even when evaluating the saliency value using (1), the time required is still of order $O(N) + O(n^2)$ [6]. Thus, the number of pixel colors should be reduced to speed up the calculation. Zhai and Shah [35] proposed a method for computing pixel-level saliency maps using only luminance. Since color information is ignored, this method has flaws. Yildirim and Ssstrunk [34] also quantized the image colors to speed up the process by performing color quantization in CIELab color space to use fewer quantization bins. Cheng et al. [6] firstly quantized each color channel into twelve different values (1728 colors) directly in RGB color space, then used a weighted average method to smooth the image, and finally computed the saliency map in the CIELab color space. In this paper, we use the same image compression approach as [6]. However, we add a gradient-minimization based smoothness algorithm to get a more homogeneous background before utilizing the compression method described in [6], and finally compute the saliency map in an opponent color space, which can achieve better results.

2.1.2 Measuring global statistics saliency

Assume that an image is available that has been pre-processed using the method described in Section 2.1.1. The color quantization in RGB color space is obtained, and the distance in an opponent color space is measured (intensity channel I, color channels RG and BY), which corresponds to the opponent theory of human perception [13]. The calculation formula for each channel is as follows where R, G and B are the RGB color space values: $I = (R + B + C)/3$; $RG = R - G$; $BY = B - (R + G)/2$.

Rather than treat all three channels I, RG, and BY equally in [8], we have found that usually only one or two channels perform well with our method for saliency computation. Furthermore, consider the three examples shown in Fig. 2, which from (a) to (h), show the source images, the I, RG and BY channel images, the corresponding saliency results of each channel and the ground truth. In the mailbox example, it can be seen that the RG channel performs the best, and its saliency map is very close to the ground truth image. The I channel is also useful,

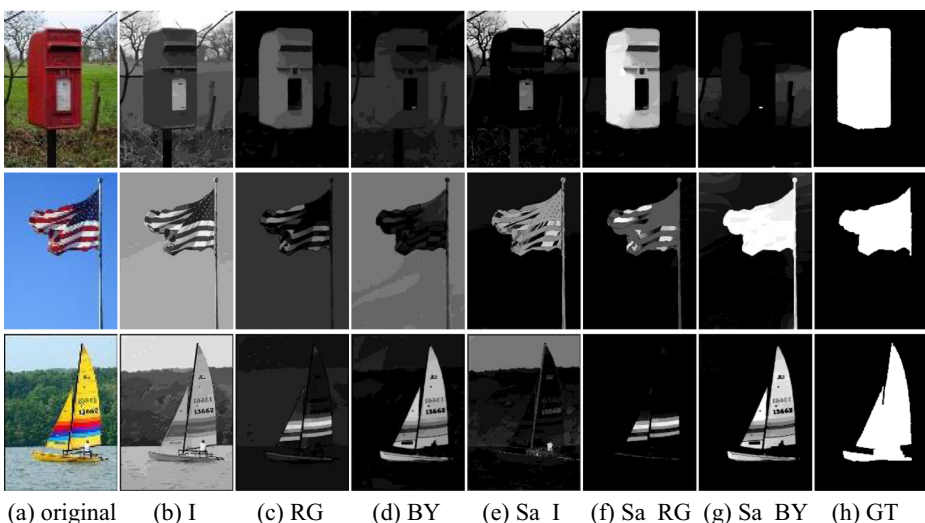


Fig. 2 Saliency computation in the opponent color space

but the BY channel does not seem to contribute much to the final saliency computation. Similar phenomena are also observed in the flag and sailboat examples. Since not every channel provides useful information for the saliency computation, conventional methods that take the average or maximum value cannot be applied. For this reason, we have proposed a weighted fusion method to fuse the saliency map of each channel based on the following simple principle: The number of salient pixels having a high color contrast with all other pixels in the image should account for only a small part of the image compared with the background area.

Based on that principle, the saliency map of each channel is firstly calculated, and then the percentage of pixels with values exceeding the average saliency values of the entire saliency map is obtained, expressed as N_I , N_{RG} , N_{BY} . A reference percentage value r is then manually set, based on empirical results. We found that a value of about 0.3 for r works well, and Eq. (3) is used to fuse the saliency maps of the three channels together:

$$S_{gs} = Norm \left(\sum (w_I \cdot S_I, w_{RG} \cdot S_{RG}, w_{BY} \cdot S_{BY}) \right) \quad (3)$$

where S_I , S_{RG} and S_{BY} respect the saliency maps of the color spaces (I, RG, BY) respectively, w_I , w_{RG} and w_{BY} are the weight coefficients of each corresponding color space, and *Norm* represents the normalization approach. Since the saliency map can easily be obtained using Eq. (1), the difficulty of the formula is only how the weight coefficient w is defined. According to the principles above, the weight coefficient should be:

$$w = \begin{cases} \min(1, 1 - (N - r)^2) & (N \leq r) \\ \max(0, 0.5 - (N - r)^{1/2}) & (\text{other}) \end{cases} \quad (4)$$

where N is the percentage of pixels that have values exceeding the average saliency value of the saliency map, and r is the reference percentage value.

In (4), if N is less than or equal to r , the weight value will be slowly reduced with r -centered, and such a measure will increase the influence of the corresponding saliency map. Conversely, if N is larger than r , this measure will weaken its influence rapidly. The resultant map is convolved with a small Gaussian kernel for final smoothing to achieve better visualization.

2.2 Surrounding contrast saliency generation

Although the proposed global statistics contrast saliency computing method can get pixel-wise saliency values, and produce full-resolution saliency maps, its main shortcoming is that it ignores the spatial relationships which are important in human attention [7]. An ideal contrast-driven saliency detection method should take both local perspective and global-homogeneous properties into account [33]. Based on that, we also proposed a surrounding contrast based saliency detection algorithm considering both color and the textural distinctness of a region with respect to its surroundings.

In contrast with many region segmentation approaches [6, 9], we use a SLIC method [3] to segment the input image into multiple regions, which are considered as basic units instead of pixels. The SLIC algorithm adopts k-means clustering to generate superpixels, which shows good performance on many widely-used datasets. After the segmentation step, a saliency computational method is proposed which exploits the strength of both the color and textural features.

2.2.1 Color distinctness

Color is an important feature, and is used in almost all saliency models. In this paper, we detect the color distinctness of patches in the CIE Lab color space, since it has high efficiency for saliency detection [6, 9]. A patch is considered to be salient if it is consistently different from other patches. We know that spatial relationships also play an important role in the measure of saliency. Furthermore, patches close to the current location will have more influence than those further away from the current patch [33].

Based on the above analysis, we specifically define the distance between patches i and j in the joint space position and color as

$$d(i, j) = \frac{d_{color}(i, j)}{1 + \alpha \cdot d_{position}(i, j)} \quad (5)$$

where $d_{color}(i, j)$ is the Euclidean distance, normalized to a range $[0, 1]$, and $d_{position}(i, j)$ is the normalized spatial distance between i and j . α is used to control the color/spatial weight proportions, and in our implementation is set at $\alpha = 1$ in our implementation. Finally, the color saliency of patch i in our model can be expressed as

$$S_{ic}(i) = Norm \left(\sum_{p_j \in N_k} d(i, j) \right) \quad (6)$$

where N_k contains the k -nearest neighbors for current patch i in terms of color distance, and $Norm$ represents the normalization approach.

2.2.2 Textural distinctness

Many different aspects of distinctness have already been examined previously. Since some regions of distinct color may be non-salient, consideration of color distinctness in isolation would be insufficient.

In this paper, more accuracy is achieved by using LBP features to determine the textural distinctness. The average value of each region for LBP features is considered using the same method as [27], and the normalized histograms for each superpixel are then calculated. i.e., a vector of 59 dimensions ($\{h_i, i=1, 2 \dots 59$, where h_i is the i -th bin in an LBP histogram}). Furthermore, the textural distinctness of a region is computed by SLIC, as the sum of L_2 distances from all other N_k regions. Given M regions, the textural distinctness of region p_i can be computed by:

$$S_{it}(p_i) = \sum_{p_j \in N_k} \|p_i - p_j\|_2 \quad (7)$$

where N_k contains the k -nearest neighbors for current patch p_i .

Additionally, the accuracy of the saliency map also relates to the number of regions. In this paper, different-scale regions are simply obtained by generating four layers of superpixels with different granularities, where $N=100, 150, 200, 250$ respectively, and each saliency map is then averaged and normalized within the range $[0, 1]$. The CA method [9] considered the K most similar patches to compute the saliency values. However, since the limited quantity of superpixels, we calculate a region's saliency by measuring its contrast to all other regions, that is, we use N_k equal to the number of regions in (6) and (7) in our experiments.

2.2.3 Measuring surrounding contrast saliency

Since it is required to find regions that are salient in both color and texture, the color and textural distinctness are combined into a saliency map by simply taking the product of the two:

$$S_{sc} = S_{lc} + \gamma \cdot S_{lt} \quad (8)$$

where S_{lc} represents the color saliency value of the input image, and S_{lt} denotes the textural saliency value. γ denotes the strength of textural distinctness weighting, which is set as $\gamma = 0.5$ in our implementation. Finally, the surrounding contrast saliency map is normalized within the range [0 1].

2.3 Final saliency map generation

Thus far, two saliency maps for the input image have been generated: the global statistics map S_{gs} and the surrounding contrast map S_{sc} . Our final saliency map is built by integrating the two maps together.

Unlike many other methods such as GB [11] or CA [9], the method proposed in this paper does not introduce any location prior mechanism, although these approaches can improve the system performance in some datasets. The reason for this is that for applications on mobile systems such as robots, objects can have any position. Thus, a location prior can do more harm than good. Since taking both the global and local saliency operators into account works better than using either individual method [4], we obtain the final saliency map S_f as

$$S_f = 0.5 * G(S_{gs}, S_{sc}) \quad (9)$$

where S_{gs} is obtained using the method in Section 2.1, S_{sc} is obtained using the method in Section 2.2, and G is a fusion operation. There are many methods that can then be used for integration of the two maps (i.e. $\{+, *, \max, \text{or}, \min\}$). Through experimentation, we have found that addition operation at this stage leads to the best results, which appears to be a similar conclusion as that reached in [28].

3 Empirical evaluation

In this section, we evaluate the proposed method compared with fourteen state-of-the-art methods, for three of the most widely-used datasets.

3.1 Saliency dataset

The first dataset is the ASD dataset [2] of 1000 images, which includes a more refined manually-segmented ground truth. The second dataset is the MSRA [20], which has 5000 images and includes accurate human-marked object-contour ground truths, which also contains the ASD dataset. The third dataset is the THUS [6], which contains 10,000 images with labeled pixel-wise ground truths, and this dataset not only includes the ASD dataset and the MSRA dataset, but also has more images than commonly-used saliency datasets.

3.2 Performance evaluation

In our experiment, we compare our method with fourteen saliency region detection methods. These are: IT [14], GB [11], FT [2], AC [1], CA [9], HC [6], RC [6], CB [15], AIM [5], SEG

[24], SUN [36], Tong et al. [28], PCA [23], Ye et al. [32]. We choose these methods due to their number of citations (IT, GB, FT, AC, CB, AIM, SUN), surrounding contrast methods (IT, GB, AC, CA, SEG, RC), global statistics methods (FT, HC) and integrated approaches (Tong, Ye). In our comparison experiments, all the cited codes were downloaded from the websites of the authors, and most of the URLs of these compared models can be found in (*mmcheng.net*).

3.2.1 Quantitative evaluation

Firstly, a saliency map was computed for each image in the test dataset and the saliency value was normalized between [0 255]. A segmentation was then generated by simply setting the threshold from 0 to 255 for the saliency map to obtain 256 binary masks where the pixels with saliency values below a given threshold were masked out. The precision and recall rate were then computed using the following definitions:

$$precision = |SF \cap GF| / |SF|, \quad recall = |SF \cap GF| / |GF|, \quad (10)$$

where SF denotes the segmented salient pixels, GF denotes the ground truth salient pixels, and $|*|$ denotes the number of pixels in a set. Finally, the precision-recall curves were computed by adjusting the threshold from 0 to 255. As shown in Figs. 3, 4 and 5, the PR curve results demonstrate that the proposed saliency method achieves the highest performance for the larger datasets (MSRA and THUS). Our method also obtains the highest precision value of 98.53 % on the ASD database.

Furthermore, the area under the true positives (TP) and false positives (FP) curve was also calculated as an AUC score. Although the AUC score is sensitive to blurring, it will still show

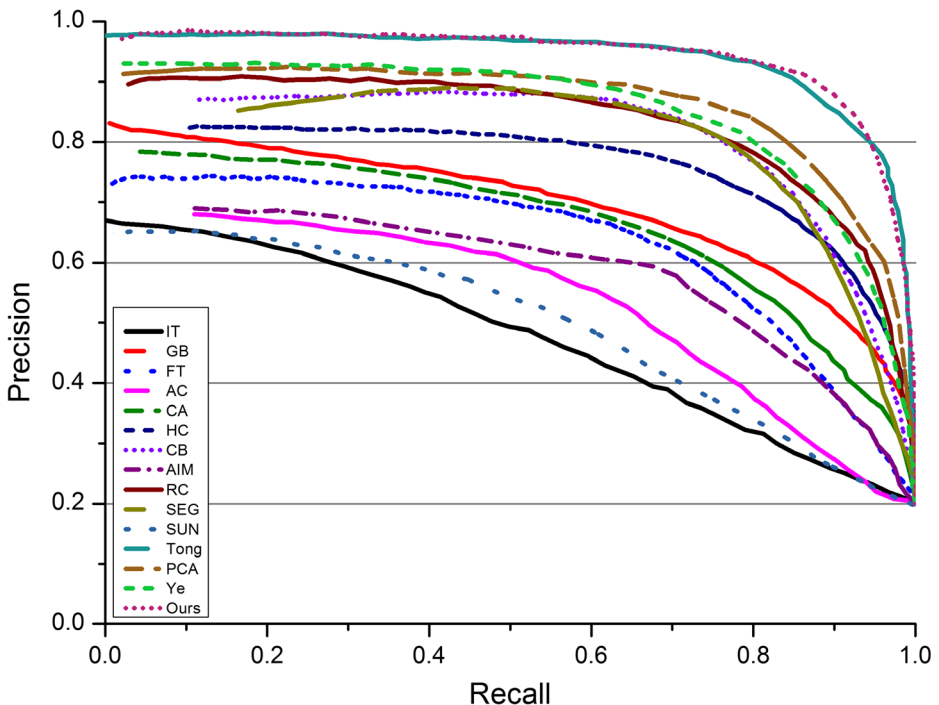


Fig. 3 PR curves of different models on ASD dataset

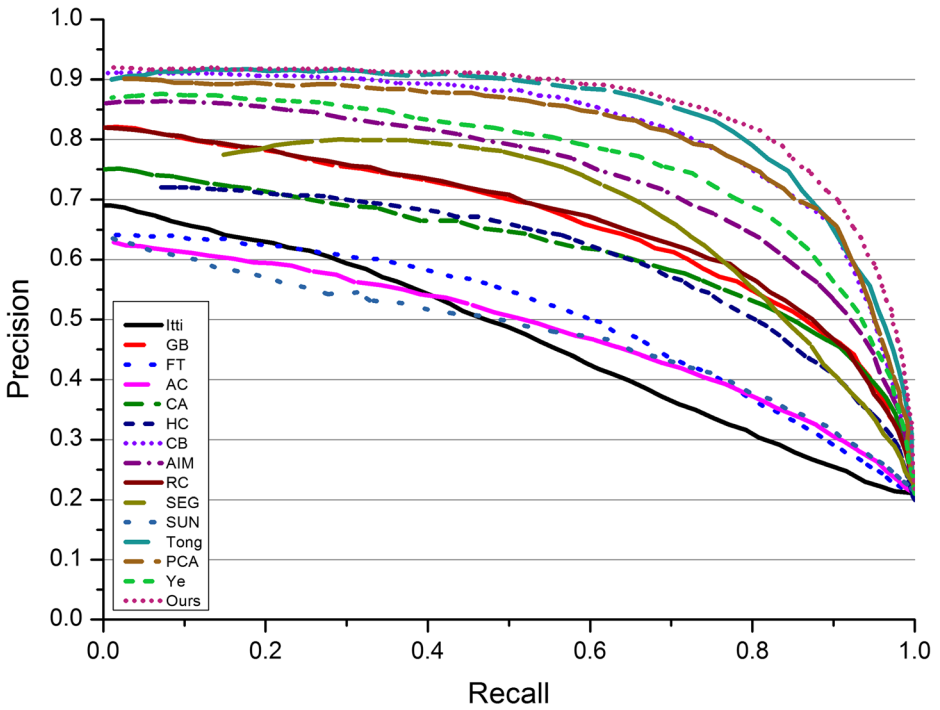


Fig. 4 PR curves of different models on MSRA dataset

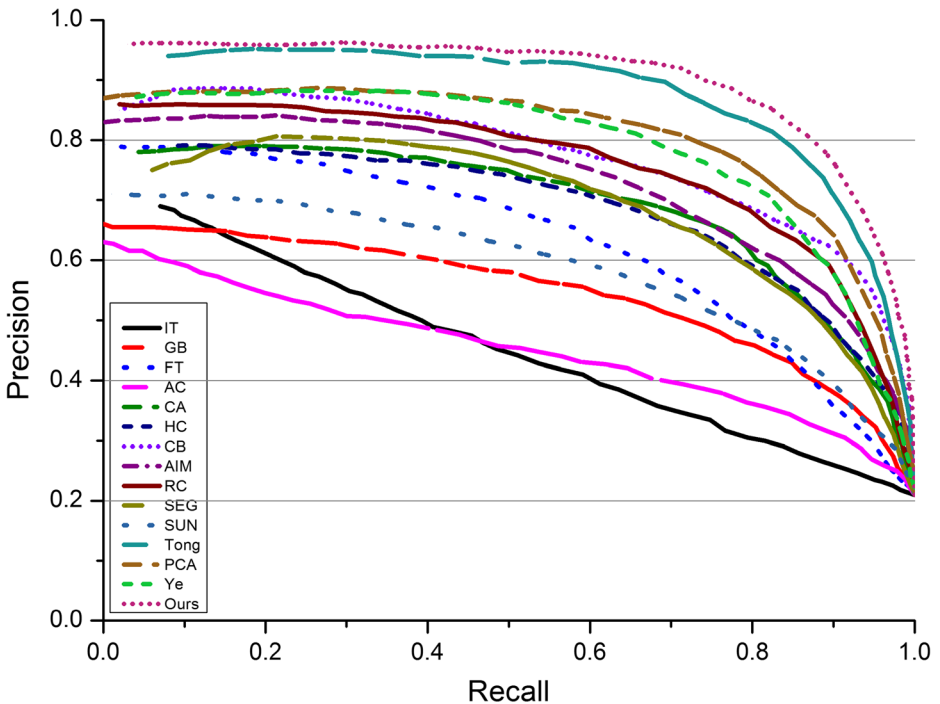


Fig. 5 PR curves of different models on MSRA dataset

the differences between each of the methods clearly and reliably. As shown in Fig. 6, the proposed method achieves the best results on all the test datasets. More specifically, the average AUC score of our model is 0.9671, which is an improvement of 0.0055 compared with the second best algorithm, and 0.0244 compared with the third best algorithm on the three datasets. Even though no location prior mechanism is introduced, our method still exceeds other methods in both AUC scores and PR curve results, largely due to the combination of global statistics and surrounding contrast information.

3.2.2 Qualitative evaluation

Some results of the saliency maps generated by the fifteen methods for qualitative comparison are presented in Fig. 7. As can be seen, the classic implementation methods of computational visual attention (e.g., IT [14] and GB [11]) generate salient regions that have low resolution and poorly defined borders. The frequency-tuned saliency detection approach (FT) [2] creates full resolution saliency maps with well-defined boundaries of salient objects, but no significant luminance difference between the salient and non-salient regions can be seen, as shown in the fourth column of Fig. 7. HC [6] can highlight the whole saliency region because the global statistics over the whole image are taken into account. Similarly, due to the integration of spatial information, RC [6] has uniformly-highlighted salient regions, as shown in the seventh and tenth columns of Fig. 7. However because these methods, including AC [1], SEG [24], Ye et al. [32], do not consider the pattern features, they cannot detect salient objects accurately when they have a similar appearance to the background regions. Additionally, CB [15] uses shape information to better define a salient object and achieves good results. However, some partial results of CB contain part of the background (e.g., the stop sign), and also the AIM [5] saliency model has the same characteristics. The SUN [36] proposes a Bayesian framework to compute the saliency map, but its performance decreased significantly when faced with a complicated background (e.g., the dog and sleigh image). The CA [9] method tended to produce higher saliency values near edges as shown in the sixth column of Fig. 7. Both our approach and Principal Component Analysis (PCA) [23] method consider the relationship between color and pattern distinctness, but due to integrating the global statistics information,

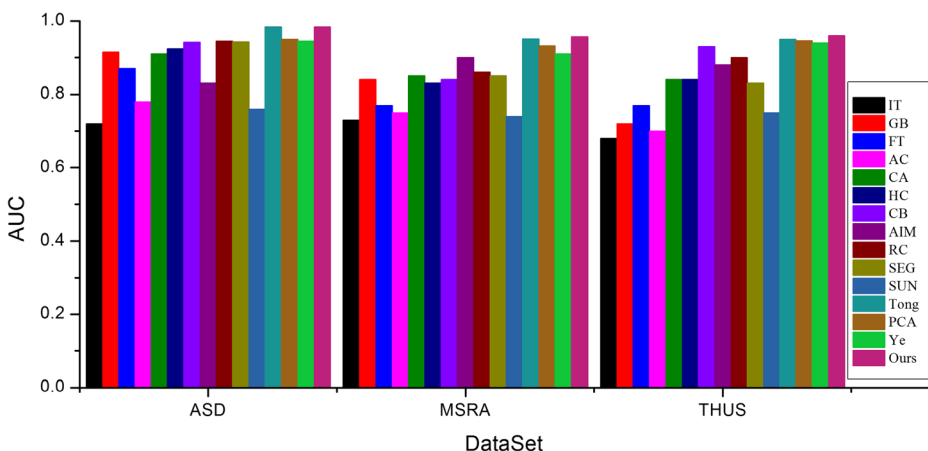


Fig. 6 Area under ROC Curve (AUC) on the ASD, MSRA and THUS datasets

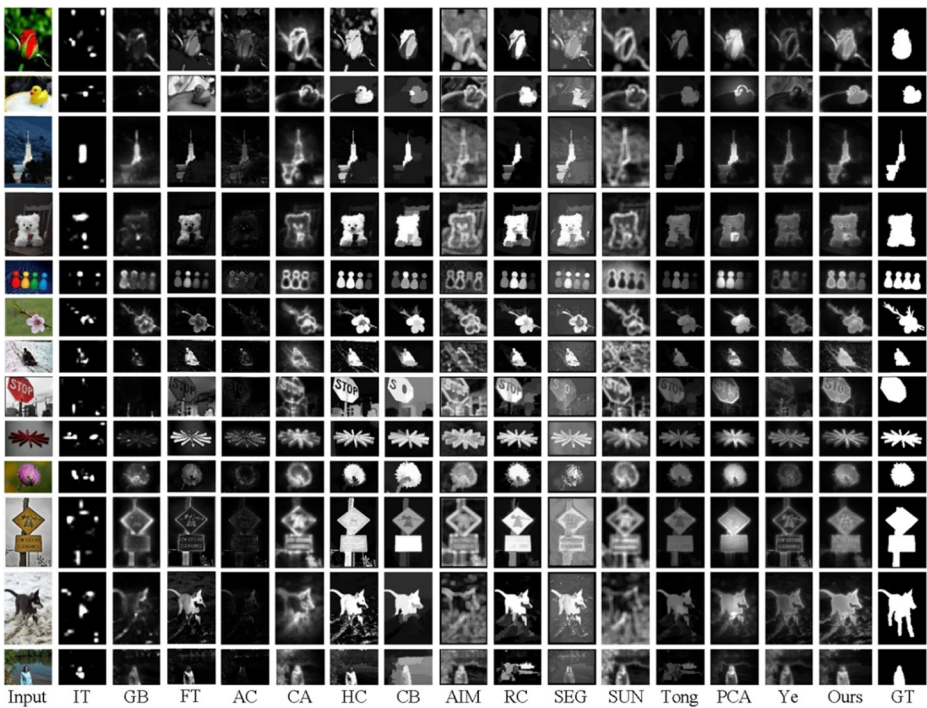


Fig. 7 Visual comparison of previous approaches to our method and ground truth

our method can generate a sparser and more accurate saliency map (e.g., the duck toy). Tong et al. [28] and Ye et al. [32] are types of integrated approaches, so they can highlight salient regions and dim the background, but some shortcomings can still be clearly seen (e.g., the stop sign and the environmental portrait). Our method integrates global statistics and surrounding contrast distinctness, and hence effectively detects both of the outlines, as well as the inner pixels of the salient region.

4 Conclusion and future works

In this paper, a new method of salient region detection has been constructed based on global statistics and surrounding contrast information. The global statistics saliency map has been constructed based on global contrast in an opponent color space. For the surrounding contrast model, a widely-used superpixel method has been used to over-segment the input image into small regions, and then the saliency values have been calculated taking the color, spatial and textural distinctness factors into account. The final saliency map has been obtained by integrating the two saliency maps, and the experimental results of fifteen state-of-the-art methods (including our method) have been compared for three datasets, and have shown that our method can achieve significantly better saliency results in quantitative analysis. In future work, we will propose some further modifications to improve the optimized performance, and focus more attention on specific applications of our algorithm, such as robot navigation and localization, path planning, motion control, etc.

References

1. Achanta R, Estrada F, Wils P, Susstrunk S (2008) Salient region detection and segmentation. In: The 6th International conference on computer vision systems, p 66–75
2. Achanta R, Hemami S, Estrada F, Susstrunk S (2009) Frequency-tuned salient region detection. In: IEEE conference on computer vision and pattern recognition, p 1597–1604
3. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Susstrunk S (2012) SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans Pattern Anal Mach Intell* 34(11):2274–2282
4. Borji A, Itti L (2012) Exploiting local and global patch rarities for saliency detection. In: IEEE conference on computer vision and pattern recognition, p 478–485
5. Bruce ND, Tsotsos JK (2009) Saliency, attention, and visual search: an information theoretic approach. *J Vis* 9(3):5
6. Cheng M, Mitra NJ, Huang X, Torr PH, Hu S (2015) Global contrast based salient region detection. *IEEE Trans Pattern Anal Mach Intell* 37(3):569–582
7. Einhauser W, König P (2003) Does luminance-contrast contribute to a saliency map for overt visual attention? *Eur J Neurosci* 17(5):1089–1097
8. Frintrop S, Werner T, Martin Garcia G (2015) Traditional saliency reloaded: a good old model in new shape. In: IEEE conference on computer vision and pattern recognition, p 82–90
9. Goferman S, Manor L, Tal A (2012) Context-aware saliency detection. *IEEE Trans Pattern Anal Mach Intell* 34(10):1915–1926
10. Hadizadeh H, Bajic IV (2014) Saliency-aware video compression. *IEEE Trans Image Process* 23(1):19–33
11. Harel J, Koch C, Perona PI (2007) Graph-based visual saliency. In: *Advances in neural information processing systems*, p 545–552
12. Hou X, Zhang L (2007) Saliency detection: a spectral residual approach. In: IEEE conference on computer vision and pattern recognition, p 1–8
13. Hurvich LM, Jameson D (1957) An opponent-process theory of color vision. *Psychol Rev* 64(6):384–404
14. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans Pattern Anal Mach Intell* 11:1254–1259
15. Jiang H, Wang J, Yuan Z, Liu T, Zheng N, Li S (2011) Automatic salient object segmentation based on context and shape prior. In: *BMVC*, p 1–12
16. Jiang H, Wang J, Yuan Z, Wu Y, Zheng N, Li S (2013) Salient object detection: a discriminative regional feature integration approach. In: IEEE conference on computer vision and pattern recognition, p 2083–2090
17. Kang D, Lee S, Lee Y-B (2011) Human visual attention with context-specific top-down saliency. In: *IEEE International Conference on Robotics and Biomimetics*, p 2055–2060
18. Li J, Levine MD, An X, Xu X, He H (2013) Visual saliency based on scale-space analysis in the frequency domain. *IEEE Trans Pattern Anal Mach Intell* 35(4):996–1010
19. Li Q, Zhou Y, Yang J (2011) Saliency based image segmentation. In: *International Conference on Multimedia Technology*, p 5068–5071
20. Liu T, Yuan Z, Sun J, Wang J, Zheng N, Tang X, Shum H-Y (2011) Learning to detect a salient object. *IEEE Trans Pattern Anal Mach Intell* 33(2):353–367
21. Luo W, Li H, Liu G, Ngan KN (2012) Global salient information maximization for saliency detection. *Signal Process Image Commun* 27(3):238–248
22. Ma YF, Zhang HJ (2003) Contrast-based image attention analysis by using fuzzy growing. In: *Proceedings of the eleventh ACM international conference on Multimedia*, p 374–381
23. Margolin R, Tal A, Zelnik-Manor L (2013) What makes a patch distinct? In: IEEE conference on computer vision and pattern recognition, p 1139–1146
24. Rahtu E, Kannala J, Salo M, Heikkilä J (2010) Segmenting salient objects from images and videos. In: *Proceedings of the European conference on computer vision*, p 366–379
25. Reynolds JH, Desimone R (2003) Interacting roles of attention and visual salience in V4. *Neuron* 37(5):853–863
26. Sharma G, Jurie F, Schmid C (2012) Discriminative spatial saliency for image classification. In: IEEE conference on computer vision and pattern recognition, p 3506–3513
27. Tong N, Lu H, Ruan X, Yang M-H (2015) Salient object detection via bootstrap learning. In: IEEE conference on computer vision and pattern recognition, p 1884–1892
28. Tong N, Lu H, Zhang Y, Ruan X (2014) Salient object detection via global and local cues. *Pattern Recogn* 48(10):3258–3267
29. Wan SH, Jin PG, Yue LH (2009) An approach for image retrieval based on visual saliency. In: *International conference on image analysis and signal processing*, p 172–175
30. Xu L, Lu C, Xu Y, Jia J (2011) Image smoothing via L0 gradient minimization. *ACM Trans Graph* 30(6)
31. Yang J, Yang M-H (2012) Top-down visual saliency via joint crf and dictionary learning. In: *International conference on image analysis and signal processing*, p 2296–2303

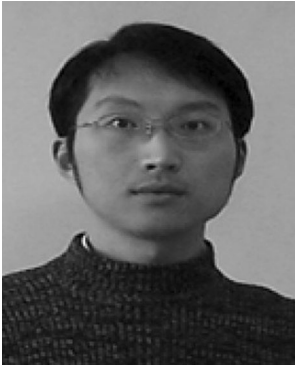
32. Ye T, Zhang D, Gao K, Jin G, Zhang Y, Yuan Q (2014) Salient region detection: Integrate both global and local cues. In: IEEE International Conference on Multimedia and Expo, p 1–6
33. Yeh H-H, Liu K-H, Chen C-S (2014) Salient object detection via local saliency estimation and global homogeneity refinement. *Pattern Recogn* 47(4):1740–1750
34. Yildirim G, Süsstrunk S (2015) FASA: fast, accurate, and size-aware salient object detection. In: Asian Conference on Computer Vision, p 514–528
35. Zhai Y, Shah M (2006) Visual attention detection in video sequences using spatiotemporal cues. In: Proceedings of the 14th annual ACM international conference on Multimedia, p 815–824
36. Zhang L, Tong MH, Marks TK, Shan H, Cottrell GW (2008) SUN: a Bayesian framework for saliency using natural statistics. *J Vis* 8(7):32
37. Zhang W, Xiong Q, Shi W, Chen S (2015) Region saliency detection via multi-feature on absorbing Markov chain. *Vis Comput*:1–13



Jie Niu currently is a Ph.D candidate in the School of Mechanical Engineering, Nanjing University of Science and Technology, NanJing, China. He is also a lecturer at School of Electrical and Electronic Engineering, Changzhou College of Information Technology, ChangZhou, China. His research interests are visual attention, visual saliency detection, machine learning and intelligent robotic system.



Xiongzhu Bu currently works in Nanjing University of Science and Technology (NUST), NanJing, China. He is currently a professor and department chairman in the department of Instrument science and technology. His current research interests are in the areas of pattern recognition, image processing and intelligent testing system.



Kun Qian received the Ph.D. in control theory and control engineering from Southeast University in 2010. He is currently working as a lecturer at School of Automation, Southeast University. His research interests are image processing, intelligent robotic system and robot control.