

Color boosted visual saliency detection and its application to image classification

Bing Yang · Duanqing Xu

Published online: 19 June 2012
© Springer Science+Business Media, LLC 2012

Abstract For many applications in graphics, design and human computer interaction, it is essential to reliably estimate the visual saliency of images. In this paper, we propose a visual saliency detection method that combines the respective merits of color saliency boosting and global region based contrast schemes to achieve more accurate saliency maps. Our method is compared with existing saliency detection methods when evaluated using four public available datasets. Experimental results show that our method consistently outperformed current state-of-the-art methods on predicting human fixations. We also demonstrate how the extracted saliency map can be used for image classification.

Keywords Visual saliency detection · Color boosted algorithm · Regional contrast · Image classification

1 Introduction

Attempting to search for and recognize particular known object in a scene can be extremely complex when one has to consider all possible views an object can take. Human vision system employs attention to try to limit the amount of information that needs to be processed in order to speed up search and recognition, and to interpret complex scenes in real time [46]. Predicting locations at where people are likely to look has many real-world applications. Computational models can be applied to various computer vision tasks such as navigational assistance [11], robot control [6], surveillance systems [18], object detection and recognition [26], and scene understanding [23]. Such predictions also find applications in other areas including adaptive image and video compression [10], pictorial database querying [7] and content-aware image editing [41].

B. Yang · D. Xu (✉)
Computer Science College, Zhejiang University, Hangzhou 310027, China
e-mail: xdq@cs.zju.edu.cn

B. Yang
e-mail: ybily061821@126.com

Visual saliency originates from visual uniqueness, unpredictability, or rarity, and is often contributed to variations in image features like color, gradient, edges and boundaries. Although many factors may determine what image features are selected or discarded by our attentional processes, visual saliency can generally be categorized into two groups: bottom-up and top-down [9]. The former comprises data driven (instantaneous) processes while the later comprises processes that are dependent of the organism's internal state (such as the visual task at hand or the subject's background) [29].

Because bottom-up saliency is important for many practical applications, we focus on bottom-up data driven saliency in this paper. The fast, parallel, pre-attentive, bottom-up stage of human vision is thought to guide a serial (computationally intensive) attentive, top-down stage. Among all features that contribute to image saliency, orientation and color are thought to be the most significant ones [22, 40]. Consequently, most current computational saliency models are based on color or orientation contrast (e.g. [8, 23, 39]). However, in color science, chromaticity determines the quality of a color regardless of its luminance, and it was reported in [35] that humans were more sensitive to chromaticity cues than to luminance ones. Moreover, there exists evidence that the human visual system combines low-level features in an early stage [2], and information theory can be used to compute the saliency by combining chromaticity and contrast in an early stage.

Therefore, in this paper, we propose a method which computes image saliency from the color saliency boosting version of original image. The method is based on the key observation that in natural images, color transitions of equal probability (i.e. isosalient transitions) form ellipsoids in decorrelated color spaces [36]. The transformation that turns these ellipsoidal surfaces into sphere ones (called the color saliency boosting function), ensures that vectors of equal length have equal information content and thus equal impact on the saliency function. In [36], the color boosted saliency map was computed by a simple gradient-based scheme. In this paper, we further investigate how to achieve a more accurate color boosted saliency map to predict human fixations.

It is widely believed that human cortical cells may be hard wired to preferentially respond to high contrast stimulus in their receptive fields [31]. Computational saliency models have been shown to successfully model human saliency by determining the contrast of image regions to their surroundings [44], using feature attributes such as intensity, color [8], and edges [1]. Inspired by currently released global contrast based visual saliency detection method [8], we utilize region based dissimilarity to produce the final color boosted saliency map. Similar to [8], the saliency value of a region is now calculated using a global contrast score, measured by the region's contrast and spatial distance to other region in the image. The main difference between our method and global regional contrast based visual saliency detection method [8] is that, we use the color boosted image instead of original image as an input. In addition, for purpose of further improvement, we initialize the segmentation using a modified version of graph-based image segmentation method [45], and contrive a spatial weighting term to describe a bias to the center of image.

The main contribution of this paper is to combine the respective merits of color saliency boosting and global region based contrast schemes to achieve more accurate saliency maps. We also have extensively evaluated our method on four publicly available benchmark datasets provided by Bruce et al. [5], Judd et al. [24], Linde et al. [37] and Achanta et al. [1], and compared our method with eight state-of-the art saliency methods [1, 5, 8, 12, 19, 21, 23, 44]. The experiments showed consistent and improvements over previous methods on all datasets, indicating the robustness and generality of proposed method.

The rest of the paper is organized as follows. Section 2 introduces the related works. Section 3 outlines the color saliency boosting algorithm we used in this paper. The regional

contrast based saliency detecting approach is presented in Section 4, and the comprehensive experimental results are given in Section 5. Section 6 demonstrates how the extracted saliency map can be used for image classification. Finally, the conclusions are outlined in Section 7.

2 Related work

2.1 Color distinctiveness

Information theory has been successfully applied in modeling human visual from image feature [25]. The theory declares that feature saliency is inversely related to feature occurrence, i.e. rare features are more informative and thereby more salient than features that occur more frequently. Consequently, recent models for predicting human visual fixation behavior suppose that saliency driven free viewing corresponds to maximizing information sampling [14]. Such models have been successfully applied to model fixation behavior, saliency asymmetries, and even to solve the classic computer vision problem such as dynamic background subtraction [14].

Based on information context of color image derivatives, van de Weijer et al. [36] proposed a so called color saliency boosting algorithm to exploit color distinctiveness of image. The key observation behind their method was that in natural images, color transitions of equal probability (i.e. isosalient transitions) form ellipsoids in decorrelated color spaces. In their paper [36], color saliency boosting was subtly designed as a generic method that can be easily adaptable to existing feature detectors. As a result, it had been successfully applied to image retrieval [33] and image classification [27].

2.2 Contrast based visual saliency detection

We pay attention to relevant literature aiming at parallel, pre-attentive, bottom-up saliency detection, which may be biologically motivated, or purely computational, or involve both aspects. In general, we slice such methods into two groups: local and global schemes.

Local contrast based methods measure the rarity of image regions with respect to (small) local neighborhoods. Among local contrast based models of saliency, the model of Itti et al. [23] was one of the most influential, summing the scale-space center-surround excitation responses of feature maps at different spatial frequencies and orientations and feeding the results into a neural network to output the final saliency map. Seo & Milanfar [32] used a self-resemblance mechanism to compute saliency, where a region with dissimilar curvature compared to its surrounding was considered as being highly salient. Bruce & Tsotsos [5] designated saliency at a location to be quantified by the self-information of the location with respect to its surrounding context—either localized pixel regions, or even the entire image. More recently, Goferman et al. [16] simultaneously modeled local low-level considerations (such as color and contrast), global considerations, visual organizational rules, and high level factors (e.g. human faces) to detect the important parts of the scene. However, such local contrast based methods have some obvious disadvantages: they tend to produce higher saliency values near edges instead of uniformly highlighting salient object.

Global contrast based method compute the saliency values of an image region using its contrast with respect to the entire image. This global rarity principle agrees better with human intuition and has been implemented in different ways. Frequency based methods perform analysis in the spectral domain [1]. Achanta et al. [1] proposed a frequency tuned

method that directly computed pixel saliency using a pixel’s color difference from average image color. In [28], low spatial variance of a feature was considered to indicate high saliency, and this made an overly strict assumption that the background everywhere was dissimilar to the object. Zhai and Shah [44] considered pixel saliency based on pixel’s contrast to all other pixels. However, for efficiency, they utilized only luminance information, thus ignoring distinctiveness clues in other channels. More recently, M. M. Chen et al. [8] presented a simple but efficient algorithm that defined pixel-level saliency using global contrast differences.

3 Color saliency boosting

We first provide a high level description of our model, with the implementation details given in the following sections. To exploit color distinctiveness of image, our method begins by boosting color saliency on the original input images. The corresponding saliency maps are then extracted by using of global region based contrast schemes. A flow chart describing this process is given in Fig. 1.

The color saliency boosting algorithm by van de Weijer et al. [36] has provided an efficient method to exploit the saliency of color edges based on information theory. The algorithm is inspired by the notion that a feature’s saliency reflects its information content as follows.

Consider a color image $f=(R,G,B)$. The information content, I , of first-order directional derivative f_d , according to information theory, is given by logarithm of its probability p :

$$I(f_d) = -\log(p(f_d)) \tag{1}$$

where $p(f_d)$ is the probability of the spatial derivative and it can be calculated from a large image database, e.g. the Corel database which consists of 40,000 images. Therefore, color image derivatives which are equally frequent, from now on named iso-salient derivatives, have equal information content.

In [36], the authors concluded that the color saliency boosting function $g(.)$ which maps iso-salient derivatives to equal saliency is required. It is our goal to model the surface of iso-salient derivatives with an ellipsoid. We note that the same influence of luminance changes on the chromatic changes in the image is reflected not only in the derivatives of first order, but also in higher order. Therefore, we use the color tensor [43] to estimate the parameter of the ellipsoid as:

$$G = \begin{bmatrix} R_x^2 + G_x^2 + B_x^2 & R_xR_y + G_xG_y + B_xB_y \\ R_xR_y + G_xG_y + B_xB_y & R_y^2 + G_y^2 + B_y^2 \end{bmatrix} \tag{2}$$

where R_x, R_y, G_x, G_y, B_x and B_y are first-order of R, G, B channel along horizontal or vertical direction respectively. This definition can be seen as a simple extension of the second

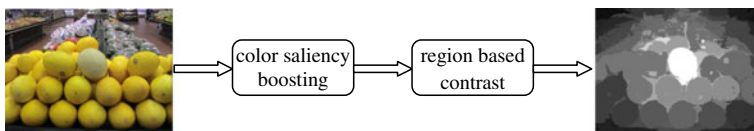


Fig. 1 A high-level overview of the proposed method



Fig. 2 Comparison for edge detection using canny detector. From top to bottom: Original images, RGB color edges, color boosted edges based on [36] and color boosted edges based on our method

moment matrix to color, and it has been successfully used to extend first order operators to color.

Applying the singular value decomposition [38], this matrix G can be decomposed into $G=U\Lambda V^T$, where U is the left singular vectors of G and then gives the color saliency boosting function g :

$$g(f) = U\Lambda^{-1/2}V^Tf \tag{3}$$

Note that due to its linearity, the color saliency boosting function g can be directly applied to the original image as a preprocessing operation, shown in Fig. 1. To evaluate the impact of our method, we conducted edge detection experiments based on three different inputs, i.e. the original image, the color boosted image using [36] and the color boosted image using our method. Figure 2 illustrates three examples from dataset [1] for edge detection using canny detector. It is obvious that RGB edges are more biased by luminance and color boosted edges based on our method are more visually apparent than edges based on [36]. Moreover, the standard quantitative evaluation is given in Table 5. Please see Table 5 for more details.

4 Region based contrast

Humans tend to focus on those image regions that contrast strongly with their neighboring regions [43]. Similar to [8], there are two factors which are considered for evaluating region based saliency: the dissimilarity between image regions, and their spatial distance. With the increasing of the spatial distance between two regions, the influence of the dissimilarity between them is decreasing. Furthermore, the distance of each region with respect to the center of an image is introduced in the spatially weighted term because of central bias [34] that was ignored in [8]. In this paper, we first segment the entire image into regions, then calculate color contrast at the region level, and measure saliency for each region as weighted sum of the region’s contrast to all other regions in the image. For each region, its corresponding spatially weighting term is contrived to include three factors: the size of this region, spatial distances between this region and all other regions, and the center bias of this region.

Due to its simplicity and efficiency, we first initialize the segmentation using the modified version of graph-based image segmentation [45] to overcome some drawbacks in [13] according to following formulas:

$$T(C) = ((W_{\max} - W_{\min})/|C|) * (Num_c/k) \tag{4}$$

$$Int(C) = 1/N * \sum_{e \in MST(C,E)} W(e) \tag{5}$$

where $T(C)$ represents threshold function, W_{\max} is the largest edge weight in the weighted graph while W_{\min} is the smallest edge weight, $|C|$ is the size of component C in pixels, Num_c is the number of components in the image, parameter k can be regarded as the expected number of components and larger k produces more components, $Int(C)$ denotes the internal difference, $MST(C,E)$ denotes the minimal spanning tree of C , $W(e)$ represents the edge weight in $MST(C, E)$ and N is the number of edges in $MST(C,E)$. We quickly review this segmentation algorithm. Initially, a graph is constructed over the entire image, with each pixel p being its own unique component. Subsequently, components are merged by traversing the edges in a sorted order by



Fig. 3 The color quantization result for an input image

increasing weight and evaluating whether the edge weight is smaller than the internal difference of both components incident to the edge. If true, the components are merged and the internal difference of the compound component is updated.

After segmentation, we then calculate region contrast by color histogram comparison. To build this color histogram while consider efficiency, we take the sparse color histogram representation proposed in [8]. In [8], each color channel was quantized into 12 different values, resulting in the number of colors into a small one $12^3=1728$. Based on the key observation that color in a natural image typically covers a small portion of the full color space, the authors in [8] further reduced the number of colors by choosing more frequently occurring colors that covered the colors of more than 95% of the image pixels, and ignoring less frequently occurring ones. The final sparse color histogram had around 85 colors, where the remaining pixels, that comprised fewer than 5% of the image pixels, were substituted by the closest colors in the histogram. Figure 3 shows that given an input image (left), we execute the color reduction procedure as discussed above. The quantized image (right) uses only 85 histogram bin colors and still remains sufficient visual quality for saliency detection.

By integrating the dissimilarity of regions *Dissimilarity*, the size of regions ω_2 , spatial distance ω_3 and central bias ω_1 , for a region r_p , we measure its saliency values to all other regions in the image as:

$$Saliency(r_p) = \omega_1(r_p) \sum_{r_p \neq r_q} \omega_2(r_q) \omega_3(r_p, r_q) Dissimilarity(r_p, r_q) \tag{6}$$

where $\omega_2(r_q)$ is the weight of region r_q and we use the number of pixels in r_q as r_q to emphasize color contrast to bigger regions. The weight $\omega_1(r_p)$ is the first weighting mechanism we proposed, indicating a bias of region r_p to the center of image, and is defined as:

$$\omega_1(r_p) = 1 - BiasToCenter(r_p)/D \tag{7}$$



Fig. 4 The impact of introducing spatial distance and central bias weighting on saliency computation

where $BiasToCenter(r_p)$ is the Euclidean distance between the centroid of region r_p and the center of image, and $D = \max_q \{BiasToCenter(r_q)\}$ is a normalization factor.

In Eq. 6, the spatial weight $\omega_3(r_p, r_q)$ is used to increase the effects of closer regions and decrease the effects of farther regions, representing the biological plausible characteristics. We calculate it as:

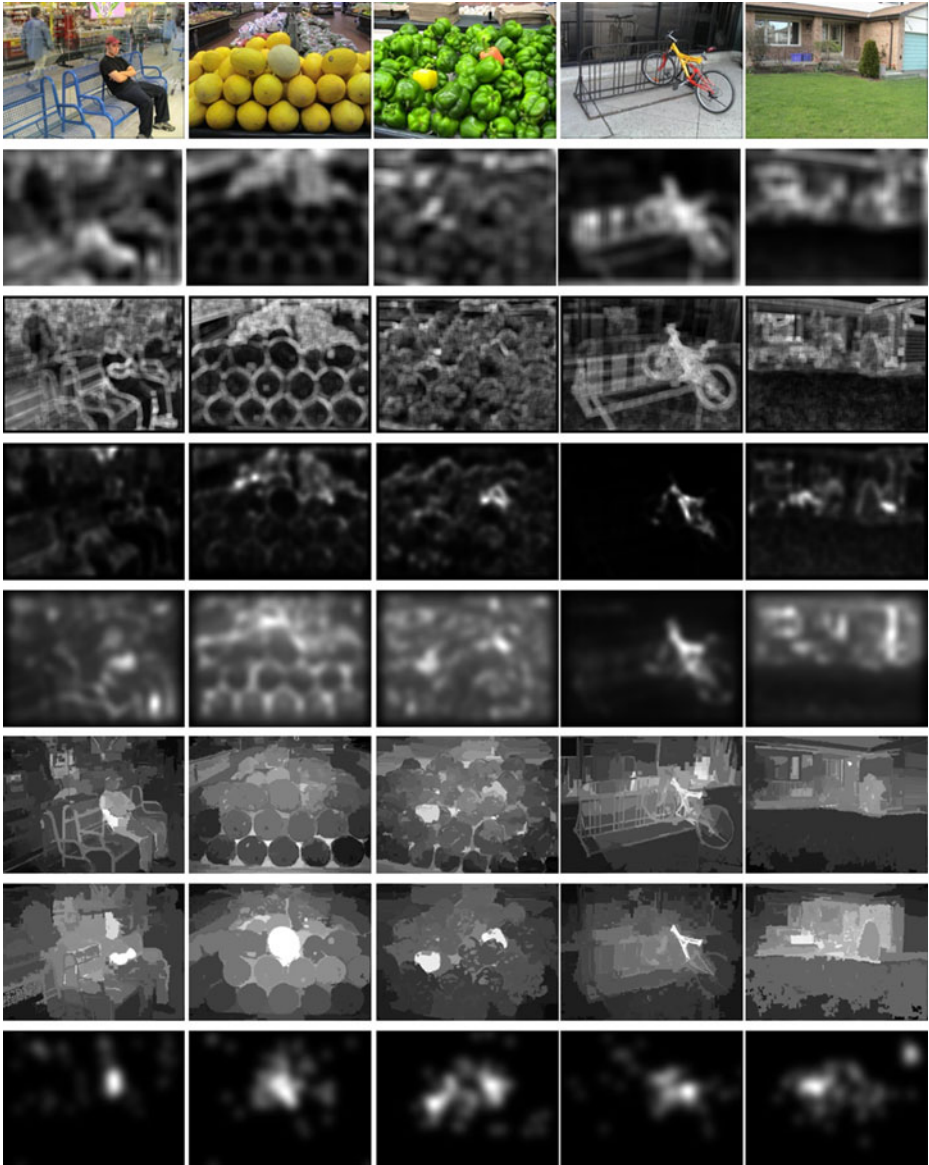


Fig. 5 Visual comparison of saliency maps between our method and other five approaches on Bruce's color image database. The rows from top to bottom are: the input images, the saliency maps of the IT method [23], IM method [5], CL method [21], GB method [19], RC method [8], our method, and human fixation density maps

Table 1 Performances on the Bruce’s color image database

Method	IT [23]	IM [5]	CL[21]	GB[19]	RC[8]	Our method
AUC	0.7092	0.7737	0.7964	0.8015	0.8300	0.8943
Improvement	–	0.0688	0.0351	0.0092	–	–

$$\omega_3(r_p, r_q) = \exp(-D_s(r_p, r_q)/\sigma_s^2) \tag{8}$$

where $D_s(r_p, r_q)$ is the spatial distance between regions r_p and r_q , computing by the Euclidean distance between their centroids, and σ_s controls the strength of spatial weighting. In fact, larger values of σ_s reduce the effect of spatial weighting so that contrast to farther regions would contribute more to the saliency of the current region. In our experiments, σ_s^2 was empirically set as 0.4.

The term *Dissimilarity* (r_p, r_q) in Eq. 6 is the dissimilarity between two regions r_p, r_q and we define it according to color distance between them as:

$$Dissimilarity(r_p, r_q) = \sum_{i=1}^m \sum_{j=1}^n f(c_{p,i})f(c_{q,j})D(c_{p,i}, c_{q,j}) \tag{9}$$

where $f(c_{p,i})$ is the probability of the i -th color $c_{p,i}$ among all m colors in the r_p region, $f(c_{q,j})$ is the probability of the j -th color $c_{q,j}$ among all n colors in the r_q region and the color distance between color $c_{p,i}$ and $c_{q,j}$ is denoted by $D(c_{p,i}, c_{q,j})$. Here, probability of a color in the probability density function (i.e. normalized color histogram) of the region is used as the weight for this color in order to emphasize more the color differences between dominant colors.

Figure 4 demonstrates the impact of introducing spatial distance ω_3 and central bias ω_1 in our method. We first use the method in [45] to segment the whole image and results are shown is the left of Fig. 4. Then, we compute the corresponding saliency map with (middle–right) and without (middle–left) spatial distance and central bias weighting. It is obvious that, by incorporating spatial distance and central bias weighting, we get a high quality saliency map when compared to the real human fixation density map (right).

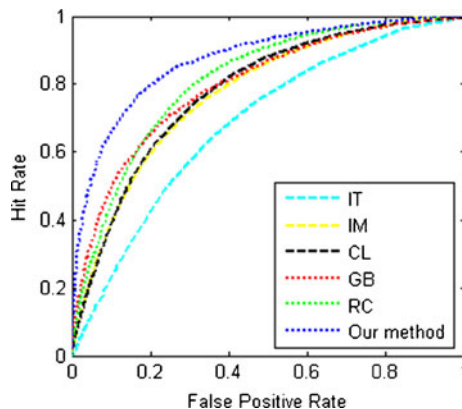


Fig. 6 The ROC curves of our method and other five approaches on Bruce’s color image database



Fig. 7 Visual comparison of saliency maps between our method and other five approaches on Judd’s color image database. The rows from top to bottom are: the input images, the saliency maps of the IT method [23], IM method [5], CL method [21], GB method [19], RC method [8], our method, and human fixation density maps

Table 2 Performances on the Judd’s color image database

Method	IT [23]	IM [5]	CL[21]	GB[19]	RC[8]	Our method
AUC	0.7615	0.7384	0.7738	0.8149	0.8170	0.8729

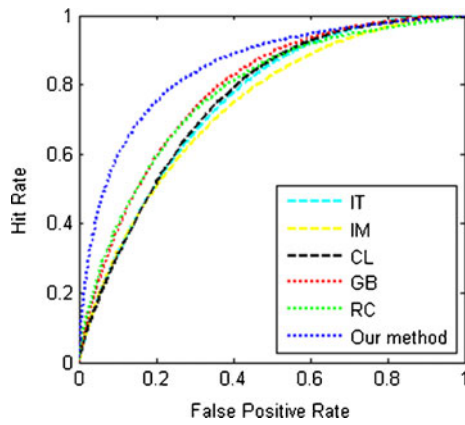


Fig. 8 The ROC curves of our method and other five approaches on Judd’s color image database

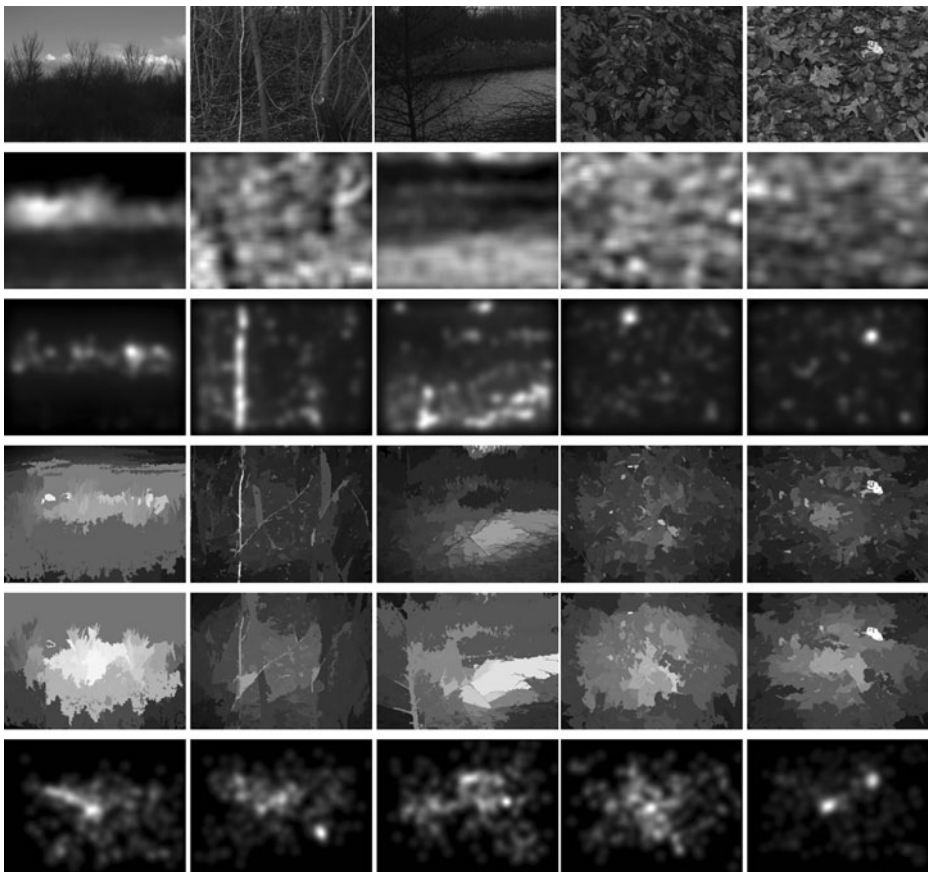


Fig. 9 Visual comparison of saliency maps between our method and other three approaches on DOVES gray image database. The rows from top to bottom are: the input images, the saliency maps of the IT method [23], GB method [19], RC method [8], our method, and human fixation density maps

Table 3 Performances on the DOVES gray image database

Method	IT [23]	GB[19]	RC[8]	Our method
AUC	0.7109	0.7941	0.8162	0.8455

5 Experimental results

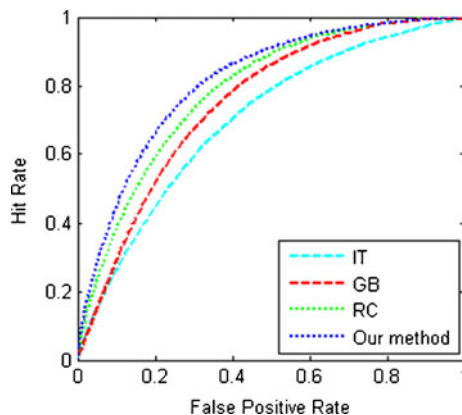
We have evaluated the results of our approach on four public databases provided by Bruce et al. [5], Judd et al. [24], Linde et al. [37] and Achanta et al. [1]. We compared the proposed method with eight state-of-the-art saliency methods, e.g. RC [8], IT [23], SR [20], FT [1], LC [44], GB [19], CL [21] and IM [5].

Please note that in our comparison experiments, the compared algorithms are different for different databases. The reasons are as follows: the authors in [1] pointed out that, “The true usefulness of a saliency map is determined by the application”. In our experiments, using the method of Receiver Operator Characteristic (ROC) curve, we found the area under the curve (AUC) of FT [1] method was smaller than that of GB [19] method. Some authors also reported the similar observation [12]. Therefore, for fair comparison, we sliced experiments into two groups: we compared our method with RC [8], IT [23], GB [19], CL [21] and IM [5] according to the AUC criterion on the three databases: Bruce, Judd and Linde, and compared our method with RC [8], SR [20], FT [1] and LC [44] on the Achanta database based on saliency-guided segmentation trials.

We implemented our method in C++. For the other methods namely, GB [19], CL [21], IM [5] and IT [23], we used the author’s implementations. In addition, we downloaded source codes, kindly provided by the authors of [8] on website, that include implementations of the RC [8], SR [20], FT [1] and LC [44] methods.

5.1 Results on Bruce’s color image database

The first color image database we used was introduced by Bruce et al. [5]. 120 different color images are contained in this database, and 20 different subjects’ fixations are recorded for each image (all the image sizes are 681×511). To objectively evaluate the performance of different methods, we utilize the area under Receiver Operator Characteristic (ROC) curve

**Fig. 10** The ROC curves of our method and other three approaches on DOVES gray image database

[5] (i.e. the area under the curve (AUC)) to compare the saliency maps with the human fixations.

Figure 5 shows visual comparison of saliency maps obtained by proposed method and five state-of-the-art methods (IT [23], IM [5], CL [21], GB [19] and RC [8]). The fixation density maps are generated from the sum of all 2D Gaussians approximations of the drop-off the density of the human fixations [5]. From Fig. 5, it is clear that the most salient locations (represented as the brightest regions) of our saliency map are more consistent with human fixation density map. For example, in the second image, the green-yellow colored melon is attended by humans, but it is not detected to be the salient region by all other saliency detection methods except ours.



Fig. 11 Visual comparison of saliency maps between our method and other four approaches on Achanta's Color Image Database. The rows from top to bottom are: the input images, the saliency maps of the SR method [20], FT method [1], LC method [44], RC method [8], our method, and manually label ground truth

Table 1 lists the results on Bruce’s color image database for six different methods. The ROC curves are generated by taking Harel et al. [19]’s code. It is clear that our method outperforms the other five methods on predicting human fixations. Note that the results of IM, CL and GB methods are different from the corresponding results published in their paper [5, 19, 21]. That is because the sampling density which we used to obtain the threshold is different from what they used. However, in our experiments, they were all evaluated according to the same validation approach, so their relative performance should not be affected.

We have also investigated the benefits of adding three weighting factors ω_1, ω_2 and ω_3 in our method. By setting ω_1 to 1, the AUC decreases from 0.8943 to 0.8726, and by setting ω_2 to 1, the AUC decreases from 0.8943 to 0.8838. Besides, if we use the flat values, i.e., ω_1, ω_2 and ω_3 are all set to 1, the AUC drastically decreases from 0.8943 to 0.7940. Therefore, all three spatially weighting factors play a significant role in computing saliency.

Furthermore, the ROC curves of six different methods on Bruce’s color image database are shown in Fig. 6 which demonstrates that our method obtains higher hit rates and lower false positive rates.

5.2 Results on Judd’s color image database

We conducted our method on Judd’s color image database [24]. There are 1003 natural images including different scenes and objects, and the corresponding human fixations are also recorded from 15 subjects. In this database, the size of the images is not the same, e.g., the width varies from 682 to 1024 pixels, and the height varies from 628 to 1024 pixels. The comparison between saliency maps using different methods is shown in Fig. 7.

The AUC results are listed in Table 2 and our method achieved highest AUC results. Furthermore, we compared the ROC curves in Fig. 8 which demonstrates that our method again obtains higher hit rates and lower false positive rates.

5.3 Results on DOVES gray image database

The DOVES gray image database we used in our experiments was introduced by van der Linde et al. [37]. This database collects eye movements from 29 human observers when they

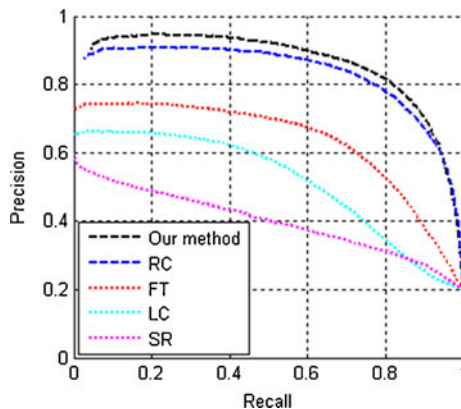


Fig. 12 Precision-recall curves for naive thresholding of saliency maps using Achanta’s color image database

Table 4 Performances on two databases with or without color saliency boosting

Color saliency boosting	Bruce's database		Judd's database	
	Yes	No	Yes	No
AUC	0.8943	0.8556	0.8729	0.8306

view 101 natural calibrated images. The database consists of around 30,000 fixation points, and is believed to be the one of the large-scale databases of eye movements to be made available to the vision research community.

Here, we compared our saliency map with that of IT [23], GB [19] and RC [8] methods, as shown in Fig. 9. Please note that IM [5] and CL [21] methods are excluded now, because they could not process gray images. Here, for gray images, the parameters hold the same values as before (e.g. the quantization in the clustering stage also takes 12 values). Based on the AUC, the qualitative results are also listed in Table 3. The corresponding ROC curves are shown in Fig. 10. Our method performs relatively poorly on DOVES gray image database. That is because the color image contains semantic objects, but the gray image contains more raw signals. However, our method is also superior to other three approaches on this database.

5.4 Results on Achanta's color image database

Finally, we evaluated the results of our method on Achanta's color image database [1]. This database contains 1003 color images, including different objects and scenes. To the best of our knowledge, the database is the largest of its kind, and has ground truth in the form of accurate human-marked labels for salient regions. Figure 11 shows saliency maps computed by our approach and other four methods (e.g. SR [20], FT [1], LC [44] and RC [8]).

To quantitatively measure the accuracy of our method on these publicly available benchmark images [1], we conducted the saliency-guided segmentation experiment. In the experiment, to segment salient regions and compute precision and recall curves [20], like the fixed thresholding experiment in [1], we binarized the saliency map using every possible fixed threshold. In other words, we vary the threshold values from 0 to 255 to reliably compare how well various saliency detection methods highlight salient regions in images. Figure 12 describes the resulting precision-recall curves generated from our method and other four methods. Our method again achieves the highest performance.

5.5 Discussion

In our proposed method, the color saliency boosting function is applied to the original image as a preprocessing step and we found it plays an important role in achieving more accurate

Table 5 Performances on two databases using different color saliency boosting functions

	Bruce's database		Judd's database	
	Our method	The method in [36]	Our method	The method in [36]
AUC	0.8943	0.8736	0.8729	0.8544

Table 6 Average time taken to compute a saliency map for images in the database by Achanta et al. [1]. All methods were tested using a Dual Core 2.4 GHz machine with 2GB RAM

Method	SR[20]	FT[1]	LC[44]	RC[8]	GB[19]	IT[23]	Our method
Time(s)	0.069	0.018	0.020	0.261	1.643	0.621	0.319
Code	Matlab	C++	C++	C++	Matlab	Matlab	C++

saliency maps. Table 4 demonstrates different performances of our method on two databases when using the color saliency boosting function or not. The performances in Table 4 are evaluated according to AUC and it is clear that with the assistant of color saliency boosting, our method could be greatly improved.

To further compare the advantage of our proposed color saliency boosting function over the method presented in [36], we conducted experiments on two databases using the AUC criterion, as shown in Table 5. Obviously, our method obtains relatively better results.

The proposed method is quite simple and efficient with no parameters are required in the main body of our method. The only parameter k , that required in Eq. 4, needs to be determined when initializing segmentation, and we set it to be 50 for all experiments.

Table 6 compares the average time taken by each method in the database by Achanta et al. [1]. Frankly speaking, our method is a little slower than four methods (e.g. SR, FT, LC and RC), but it can produce superior quality saliency maps.

6 Application: saliency based image classification

Image classification is a fundamental problem in computer vision. While steady progress has been made toward this objective, the gap between the capabilities of the primate visual system and state-of-the-art object recognition systems remains vast. In this section, we try to apply the inherent capability of human brain known as visual attention to image classification. Figure 13 describes the framework of our saliency based image classification method.

Our method is similar to NIMBLE framework [3]. For readers who are interested in such technique, please refer to the literature [3] for more details. However, most of its details, such as its features and saliency map model, have been replaced in our method. In fact, we use Independent Component Analysis (ICA) features and our proposed color boosted visual saliency detection model for image classification.

We evaluated our saliency based image classification method on the public dataset: Caltech-256 [17]. The Caltech-256 dataset [17] contains 29,780 images falling into 256 categories with much higher intra-class variability and higher object location variability. Figure 14 shows classification examples from Caltech-256 dataset.

In classification experiments, we performed 5-fold random cross validation. During per cross-validation run, for each class, n training images were randomly selected, where n was varied, and

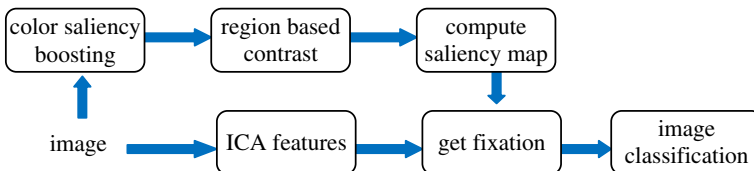


Fig. 13 The framework of saliency based image classification



Fig. 14 Classification examples from the Caltech-256 dataset

up to 30 test images randomly were chosen (distinct from the training images) unless fewer than 30 were available in which case all of the available images were used. After each cross-validation run, we computed the mean per class accuracy. We finally reported the mean accuracy and compared our results against recent state-of-the-art results [4, 15, 17, 30, 42].

Figure 15 shows our Caltech-256 results compared with other recent methods [4, 15, 17, 30, 42]. It is clear that our results are well enough compared to other approaches using a single feature type, but they are inferior to the method when multiple feature types are included. For instance, Gehler and Nowozin [15] used five feature types to train a SVM then used boosting to combine the kernels to achieve higher classification accuracy whereas our method used a single feature type.

7 Conclusion and future work

In this paper, we presented a visual saliency detection method that combines the respective merits of color saliency boosting and global region based contrast schemes to compute high accurate saliency maps. We comprehensively evaluated our method on four publicly available databases and compared our approach with eight other state-of-the-art methods. Experimental results clearly indicate the proposed method to be superior in terms of both visual comparison and quantitative analysis. Toward practical application, we also demonstrate how the extracted saliency map can be used for image classification.

There is still room for improvement. For instance, it may be useful to incorporate high level features such as human faces, symmetry into saliency map. In addition, we try to develop a more efficient saliency detection algorithm to better handle special scenes that of

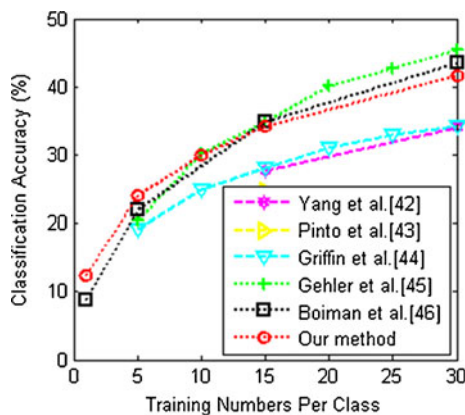


Fig. 15 The performance of our saliency based image classification on the Caltech-256 [17] dataset compared to five state-of-the-art approaches [4, 15, 17, 30, 42]

cluttered and textured background. We believe the proposed saliency detection method can be applied to many areas, e.g., efficient object detection and robust image scene analysis.

Acknowledgments This work was supported by National Program on Key Basic Research Project (973 Program), under Grant No. 2012CB725305 and by The National Key Technology R&D Program of China, under Grant No. 2012BAH03F02.

References

- Achanta R, Hemami S, Estrada F, Süsstrunk S (2009) Frequency-tuned salient region detection, in Proc. CVPR, pp. 1597–1604
- Ansar RK, Li Z (2007) Feature-specific interactions in salience from combined feature contrasts: Evidence for a bottom-up saliency map in V1. *Journal of Vision*, vol. 7, no. 7(6), pp. 1–14
- Barrington L, Marks TK, Hsiao JH-W, Cottrell GW (2008) NIMBLE: a kernel density model of saccade-based visual memory. *Journal of Vision*, vol. 8 no. 14, pp. 17: 1–14
- Boiman O, Shechtman E, Irani M (2008) In defense of nearest-neighbor based image classification, in Proc. CVPR, pp. 1–8
- Bruce NDB, Tsotsos JK (2006) Saliency based on information maximization, in Proc. NIPS, pp. 155–162
- Butko NJ, Zhang L, Cottrell GW, Movellan JR (2008) Visual saliency model for robot cameras, in Proc. ICRA, pp. 2398–2403
- Chen T, Cheng MM, Tan P, Shamir A, Hu SM (2009) Sketch2Photo: internet image montage, *ACM Trans Graph* vol. 28, no. 5(124), pp. 1–10
- Cheng MM, Zhang GX, Mitra NJ, Huang XL, Hu SM (2011) Global contrast based salient region detection, in Proc. CVPR, pp. 409–416
- Chikkerur S, Serre T, Tan C, Poggio T (2010) What and where: a Bayesian inference theory of attention. *Vision Res* 50(22):2233–2247
- Christopoulos C, Skodras A, Ebrahimi T (2000) The JPEG2000 still image coding system: an overview. *IEEE Trans Consumer Elec* 46(4):1103–1127
- Dale R, Geldof S, Prost J (2003) CORAL: using natural language generation for navigational assistance, in Proc. ACSC2003 vol. 16, pp. 1–10
- Duan L, Wu C, Miao J, Qing L, Yu Fu (2011) Visual saliency detection by spatially weighted dissimilarity, in Proc. CVPR, pp. 473–480
- Felzenszwalb PF, Huttenlocher DP (2004) Efficient graph-based image segmentation. *IJCV* 59(2):167–181
- Gao D, Mahadevan V, Vasconcelos N (2008) On the plausibility of the discriminant center-surround hypothesis for visual saliency. *Journal of Vision*, vol. 8, no. 7(13), pp. 1–18
- Gehler PV, Nowozin S (2009) On feature combination for multiclass object classification, in Proc. ICCV, pp. 221–228
- Goferman S, Zelnik-Manor L, Tal A (2010) Context-aware saliency detection, in Proc. CVPR, pp. 2376–2383
- Griffin G, Holub A, Perona P (2007) The Caltech-256, Caltech Technical Report 7694
- Guraya FFE, Cheikh FA, Tremeau A, Tong Y, Konik H (2010) Predictive saliency maps for surveillance videos, in Proc. DCABES, pp. 508–513
- Harel J, Koch C, Perona P (2006) Graph-based visual saliency, in Proc. NIPS, pp. 545–552
- Hou X, Zhang L (2007) Saliency detection: a spectral residual approach. *Proc CVPR* 1(800):1–8
- Hou X, Zhang L (2008) Dynamic visual attention: searching for coding length increments, in Proc. NIPS, pp. 681–688
- Itti L, Koch C (2001) Computational modeling of visual attention. *Nat Rev Neurosci* 2(3):194–203
- Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans Pattern Anal Mach Intell* 20(11):1254–1259
- Judd T, Ehinger K, Durand F, Torralba A (2009) Learn to predict where humans look, in Proc. ICCV, pp. 2106–2113
- Kadir T, Zisserman A, Brady M, Brady M (2004) An affine invariant salient region detector, in Proc. ECCV, pp. 228–241
- Khan FS, van de Weijer J, Vanrell M (2009) Top-down color attention for object recognition, in Proc. ICCV, pp. 979–986
- Koen van de Sande EA, Gevers T, Snoek CGM (2010) Evaluating color descriptors for object and scene recognition. *IEEE Trans Pattern Anal Mach Intell* 32(9):1582–1596

28. 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
29. Murray N, Vanrell M, Otazu X, Párraga CA (2011) Saliency estimation using a non-parametric low-level vision model, in *Proc. CVPR*, pp. 433–440
30. Pinto N, Cox D, DiCarlo J (2008) Why is real-world visual object recognition hard? *PLoS Comput Biol* 4(1):151–156
31. Reynolds J, Desimone R (2003) Interacting roles of attention and visual salience in V4. *Neuron* 37(5):853–863
32. SeoHJ, Milanfar P (2009) Static and space-time visual saliency detection by self-resemblance. *Journal of Vision*, vol. 9, no. 12(15), pp. 1–27
33. Stottinger J, Hanbury A, Gevers T, Sebe N (2009) Lonely but attractive: sparse color salient points for object retrieval and categorization, in *Proc. CVPR Workshops*, pp. 1–8
34. Tatler BW (2007) The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, vol. 7, no. 14(4), pp. 1–17
35. Tatler BW, Baddeleyb RJ, Gilchrist ID (2005) Visual correlates of fixation selection: effects of scale and time. *Vision Res* 45(5):643–659
36. van de Weijer J, Gevers T, Bagdanov AD (2006) Boosting color saliency in image feature detection. *IEEE Trans Pattern Anal Mach Intell* 28(1):150–156
37. van der Linde I, Rajashekar U, Bovik AC, Cormack LK (2009) DOVES: a database of visual eye movements. *Spat Vis* 22(2):161–177
38. Vigo DAR, van de Weijer J, Gevers T (2010) Color edge saliency boosting using natural image statistics, in *Proc. CGIV*
39. Wolfe J (2007) Guided search 4.0: current progress with a model of visual search. *Integrated models of cognitive systems*, pp. 99–119
40. Wolfe J, Horowitz T (2004) What attributes guide the deployment of visual attention and how do they do it? *Nat Rev Neurosci* 5(6):495–501
41. Wu H, Wang YS, Feng KC, Wong TT, Lee TY, Heng PA (2010) Resizing by symmetry-summarization. *ACM Trans. Graph.*, vol. 29, no. 6(159), pp. 1–9
42. Yang J, Yu K, Gong Y, Huang T (2009) Linear spatial pyramid matching using sparse coding for image classification, in *Proc. CVPR*, pp. 1794–1801
43. Zenzo SD (1986) A note on the gradient of a multi-image. *Journal of Computer Vision, Graphics, and Image Processing* 33(1):116–125
44. Zhai Y, Shah M (2006) Visual attention detection in video sequences using spatiotemporal cues, in *Proc. ACM MM*, pp. 815–824
45. Zhang M, Alhaji R (2006) Improving the graph-based image segmentation method, in *Proc. ICTAI*, pp. 1–6
46. Zhao Q, Koch C (2011) Learning a saliency map using fixated locations in natural scenes. *Journal of Vision*, vol. 11, no. 3(9), pp. 1–15



Bing Yang received her B.S. degree from Fuyang Teachers College, in 2006 and M.S. degree from Hangzhou Dianzi University, China, in 2009. She is currently working toward the Ph.D. degree in Computer Science College, Zhejiang University, Hangzhou, China. Her current research interests include pattern recognition, machine learning and aesthetic style research on Dunhuang murals.



Duanqing Xu received his Ph.D. degree in Computer Science from Zhejiang University of China in 2003. He is currently a professor at College of Computer Science, Zhejiang University, China. His main research interests include computer graphics, image processing, virtual reality and its applications.