Salient Region Detection by UFO: Uniqueness, Focusness and Objectness

Peng Jiang [1] Haibin Ling [2*] Jingyi Yu [3] Jingliang Peng [1*]
[1]School of Computer Science and Technology, Shandong University, Jinan, China
[2]Computer & Information Science Department, Temple University, Philadelphia, PA USA
[3]Department of Computer and Information Sciences, University of Delaware, Newark, DE USA
jump@mail.sdu.edu.cn, hbling@temple.edu, yu@cis.udel.edu, jpeng@sdu.edu.cn

Abstract

The goal of saliency detection is to locate important pixels or regions in an image which attract humans' visual attention the most. This is a fundamental task whose output may serve as the basis for further computer vision tasks like segmentation, resizing, tracking and so forth.

In this paper we propose a novel salient region detection algorithm by integrating three important visual cues namely uniqueness, focusness and objectness (UFO). In particular, uniqueness captures the appearance-derived visual contrast; focusness reflects the fact that salient regions are often photographed in focus; and objectness helps keep completeness of detected salient regions. While uniqueness has been used for saliency detection for long, it is new to integrate focusness and objectness for this purpose. In fact, focusness and objectness both provide important saliency information complementary of uniqueness. In our experiments using public benchmark datasets, we show that, even with a simple pixel level combination of the three components, the proposed approach yields significant improvement compared with previously reported methods.

1. Introduction

Humans have the capability to quickly prioritize external visual stimuli and localize their most interest in a scene. As such, how to simulate such human capability with a computer, *i.e.*, how to identify the most salient pixels or regions in a digital image which attract humans' first visual attention, has become an important task in computer vision. Further, results of saliency detection can be used to facilitate other computer vision tasks such as image resizing, thumbnailing, image segmentation and object detection.

Due to its importance, saliency detection has received intensive research attention resulting in many recently proposed algorithms. The majority of those algorithms are based on low-level features of the image such as appearance uniqueness in pixel or superpixel level (See Sec. 2). One basic idea is to derive the saliency value from the local contrast of various channels, such as in terms of *uniqueness* defined in [29]. While uniqueness often helps generate good saliency detection results, it sometimes produces high values for non-salient regions, especially for regions with complex structures. As a result, it is desired to integrate complementary cues to address the issue.

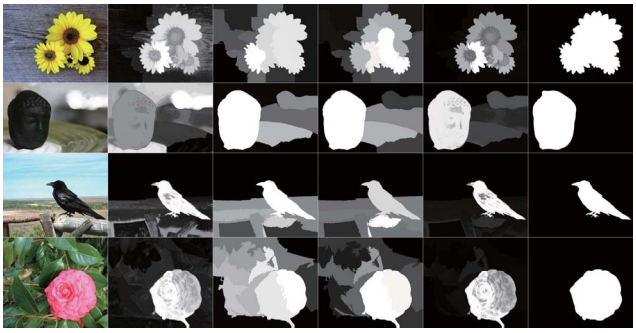


Figure 1. From left to right: source images, uniqueness, focusness, objectness, combined results and ground truth.

Inspired by the above discussion, in this paper we propose integrating two additional cues, *focusness* and *objectness* to improve salient region detection. First, it is commonly observed that objects of interest in an image are often photographed in focus. This naturally associates the focusness (*i.e.*, degree of focus) with the saliency. We derive an algorithm for focusness estimation by treating focusness as a reciprocal of blurriness, which is in turn estimated by the scale of edges using scale-space analysis. Second, intuitively, a salient region usually completes objects instead of cutting them into pieces. This suggests us to use object completeness as a cue to boost the salient region detection. The recently proposed objectness estimation method [3] serves well for this purpose by providing the likelihood that a region belongs to an object.

* Corresponding authors.

Combining focusness and objectness with uniqueness, we propose a new salient region detection algorithm, named *UFO saliency*, which naturally addresses the aforementioned issues in salient region detection. To evaluate the proposed approach, we apply it first to the intensively tested MSRA-1000 dataset [2] and then to the challenging BSD-300 dataset [25]. In both experiments, our method demonstrates excellent performance in comparison with state-of-the-arts. Finally, the source code and experimental results of the proposed approach are shared for research uses.[1]

2. Related Work

2.1. Saliency Detection

According to [26, 35], saliency can be computed either in a bottom-up fashion using low level features or in a top-down fashion driven by specific tasks.

Many early works approach the problem of saliency detection with bottom-up methods. Koch *et al*. [19] suggest that saliency is determined by center-surround contrast of low-level features. Itti *et al*. [14] define image saliency using a Difference of Gaussians approach. Motivated by this work, some approaches were proposed later which combine local, regional and global contrast-based features [1, 12, 22, 24]. Also some methods turn to the frequency domain to search for saliency cues [10, 13, 21]. The above methods strive to highlight the object boundaries without propagating saliency to the areas inside, limiting their applicability for some vision tasks like segmentation.

Later on, many works were proposed which utilize various types of features in a global scope for saliency detection. Zhai and Shah [40] compute pixel-level saliency using the luminance information. Achanta *et al*. [2] achieve globally consistent results by defining pixel's color difference from the average image color. However, these two methods do not take full advantage of color information and therefore may not give good results for images (*e.g.*, natural images) with high color complexity. Cheng *et al*. [7] study color contrast in the *Lab* color space and measure the contrast in the global scope. Perazzi *et al*. [29] promote Cheng *et al*.'s work through elements distribution analysis and propose a linear-time computation strategy. Depth cues are also introduced to saliency analysis by Niu *et al*. [27] and Lang *et al*. [23]. These methods heavily depend on color information and therefore may not work well for images with not much color variation, especially when foreground and background objects have similar colors. Comparing with these works, our study focuses more on image statistics extracted from edges.

High-level information from priors and/or special object detectors (*e.g.*, face detector) has also been incorporated into recently proposed algorithms. Wei *et al*. [37] turn to background priors to guide the saliency detection. Goferman *et al*. [11] and Judd *et al*. [18] integrate high-level information, making their methods potentially suitable for specific tasks. Shen and Wu [34] unify the higher-level priors to a low rank matrix recovery framework. As a fast evolving topic, there are many other emerging saliency detection approaches worth notice. For example, shape prior is proposed in [15], context information is exploited in [36], region-based salient object detection is introduced in [16], and manifold ranking approach is introduced for saliency detection in [39], submodular optimization-based solution is presented in [17], hierarchical saliency is exploited in [38], etc.

Borji *et al*. [4] compare the state-of-the-art algorithms on five databases. They find that combining evidences (features) from existing approaches may enhance the saliency detection accuracy. On the other hand, their experiment also shows that simple feature combination does not guarantee the improvement of saliency detection accuracy, suggesting that the widely used features may not be complementary and some may even be mutually exclusive to each other.

2.2. Uniqueness, Focusness and Objectness

In the following we briefly summarize the work related to the three ingredients used in our approach. Uniqueness stands for the color rarity of a segmented region or pixel in a certain color space. Cheng *et al*. [7] and Perazzi *et al*. [29] mainly rely on this concept to detect saliency. It is worth noting that the two methods use different segmentation methods to get superpixels (regions) and the results turn out to be very different, suggesting the important role of segmentation algorithms in saliency region detection.

We use the term focusness to indicate the degree of focus. Focusness of an object is usually inversely related to its degree of blur (blurriness) in the image. Focusness or blurriness has been used for many purposes such as depth recovery [41] and defocus magnification [31]. The blurriness is usually measured in edge regions and it is therefore a key step to propagate the blurriness information to the whole image. Bae and Durand [31] use colorization method to spread the edge blurriness which may work well only for regions with smooth interiors. Zhuo *et al*. [41] use image matting method to compute the blurriness of non-edge pixels. Baveye *et al*. [30] also compute saliency by taking blur effects into account, but their method identifies blur by wavelet analysis while our solution by scale space analysis.

The term objectness, proposed by Alexe *et al*. [3], measures the likelihood of there being a complete object around a pixel or region. The measurement is calculated by fusing hybrid low level features such as multi-scale saliency, color contrast, edge density and superpixels straddling. The objectness is later used popularly in various vision tasks such as object detection [6] and image retargeting [32].

[1] http://www.dabi.temple.edu/~hbling/code/UFO-saliency.zip

3. Salient Region Detection by UFO

3.1. Problem Formulation and Method Overview

We now formally define the problem of salient region detection studied in this paper. We denote an input color image as $I : \Lambda \to \mathbb{R}^3$, where $\Lambda \subset \mathbb{R}^2$ is the set of pixels of I. The goal is to compute a saliency map denoted as $\mathcal{S} : \Lambda \to \mathbb{R}$, such that $\mathcal{S}(\mathbf{x})$ indicates the saliency value of pixel $\mathbf{x}$.

Given the input image I, the proposed UFO saliency first calculates the three components separately, denoted as $\mathcal{U} : \Lambda \to \mathbb{R}$ for uniqueness, $\mathcal{F} : \Lambda \to \mathbb{R}$ for focusness, and $\mathcal{O} : \Lambda \to \mathbb{R}$ for objectness. The three components are then combined into the final saliency $\mathcal{S}$.

Although the saliency map is defined for per pixel, we observe that region-level estimation provides more stable results. For this purpose, in the preprocessing stage, we segment the input image into a set of non-overlapping regions, $\Lambda_i, i = 1, \ldots, N$, such that

$$\Lambda = \bigcup\nolimits_{1 \le i \le N} \Lambda_i.$$

A good segmentation for our task should reduce broken edges and generate regions with proper granularity. In our implementation we use the mean-shift algorithm [5] for this purpose.

In the following subsections we give details on how to calculate each component and hwo to combine them for the final result.

3.2. Focusness Estimation by Scale Space Analysis

Pixel-level Focusness. In general, sharp edges of an object may get spatially blurred when projected to the image plane. There are three main types of blur: *penumbral blur* at the edge of a shadow, *focal blur* due to finite depth of field and *shading blur* at the edge of a smooth object [8].

Focal blur occurs when a point is out of focus, as illustrated in Fig. 2. When the point is placed at the focus distance, d_f, from the lens, all the rays from it converge to a single sensor point and the image will appear sharp. Otherwise, when $d \neq d_f$, these rays will generate a blurred image in the sensor area. The blur pattern generated this way is called the *circle of confusion* (CoC), whose size is determined by the diameter c. The focusness can be derived from the degree of blur.

The effect of focus/defocus is often easier to be identified from edges than from object interiors. According to [8], the degree of blur can be measured by the distance between each pair of minima and maxima of the second derivative responses of the blurred edge. In practice, however, second derivatives are often sensitive to noise and clutter edges. Therefore, it is often hard to accurately localize extrema of the second derivatives [31].

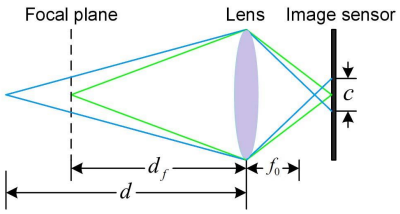


Figure 2. A thin lens model for image blur (revised from [41]).

The defocus blur can be modeled as the convolution of a sharp image [28], denoted by $E(\mathbf{x})$, with a point spread function (PSF) approximated by a Gaussian kernel $\Phi(\mathbf{x}, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp(\frac{-|\mathbf{x}|^2}{2\sigma^2})$. The scale $\sigma = \lambda c$ is proportional to the CoC diameter c, and can be used to measure the degree of blur. Consequently, the estimation of focusness relies on the estimation of the scale of edges, *i.e.*, σ.

Inspired by Lindeberg's seminal work on scale estimation [20], we derive an approach for estimating σ. In particular, let $I_E(x)$ be a 1D edge model depicting a vertical edge at position t,

$$I_E(x) = \begin{cases} k + h & \text{if } x < t; \\ k & \text{otherwise.} \end{cases}$$

The blurred edge image $I(x)$ can be modeled as the convolution of $I_E(x)$ with the Gaussian kernel, $I(x) = I_E(x) \otimes \Phi(x, \sigma)$. Denoting the Differential-of-Gaussian (DOG) operation by $\nabla_g(x, \sigma_1) = \nabla(x) \otimes \Phi(x, \sigma_1)$, the response of DOG on I is

$$f(t, \sigma_1) = \nabla_g(x, \sigma_1) \otimes \Phi(x, \sigma) \otimes I_E(x) \qquad (1)$$

within the neighborhood of an edge pixel, the response reaches its maximum when $t = 0$. Let $f(\sigma_1) = f(0, \sigma_1)$, denote the response on the edge pixel

$$f(\sigma_1) = \frac{h\sigma_1^2}{\sqrt{2\pi(\sigma^2 + \sigma_1^2)}},$$

its first and second derivatives with respect to σ_1 are

$$f'(\sigma_1) = \frac{h\sigma_1(2\sigma^2 + \sigma_1^2)}{\sqrt{2\pi}(\sigma^2 + \sigma_1^2)^{\frac{3}{2}}}, \quad f''(\sigma_1) = \frac{h\sigma^2(2\sigma^2 - \sigma_1^2)}{\sqrt{2\pi}(\sigma^2 + \sigma_1^2)^{\frac{5}{2}}},$$

It can be proven that when $\sigma_1 = \sqrt{2}\sigma$, $f''(\sigma_1) = 0$. It means that $f'(\sigma_1)$ reaches its maximum. The above derivation leads to the following way to calculate the focusness at edge pixels of an input image I,

1. Detect edges from I;
2. For each edge pixel $\mathbf{x}$, calculate its DOG responses f using different scales in $\Sigma = \{1, 2, \ldots, 16\}$;
3. Estimate f' at $\mathbf{x}$ as $f' = (f(j) - f(j-1) : j = 2, \ldots, 16)$;

4. Define the degree of blur $b(\mathbf{x})$ at $\mathbf{x}$ as $b(\mathbf{x}) = \frac{\sqrt{2}}{2} \arg\max_j(f')$;
5. Approximate the pixel-level focusness of $\mathbf{x}$ as $\mathcal{F}_p(\mathbf{x}) = \frac{1}{b(\mathbf{x})}$.

In our implementation, we set half of the window width of the filters $w = 4\sigma_1$, since $2\sigma_1$ corresponds to the distance between the peak and valley of the DOG filter and $[-w, w]$ thus covers the dominant part of the filter.

Region-level Focusness. It would be ideal to compute the saliency for each object as a whole. However, accurate object segmentation by itself is a hard problem and we hence make saliency computation in the sub-object level instead. Specifically, we conduct saliency computation for each separate region Λ_i, $i = 1, \ldots, N$.

For region Λ_i, we use B_i to denote the set of m_i boundary pixels, and E_i to denote the set of n_i interior edge pixels. It naturally follows that the focusness of Λ_i is positively related to the sum of the focusness values at all the pixels in $B_i \cup E_i$. Further, observing that a region with a sharper boundary usually stands out more salient, we use the boundary sharpness as a weight in the computation. The boundary sharpness is quantified as the mean of the gradient values, as obtained with the DOG operator, at the boundary pixels. Specifically, we formulate the region-level focusness, $\mathcal{F}_r(\Lambda_i)$, of Λ_i as:

$$\mathcal{F}_r(\Lambda_i) = \frac{1}{m_i} \sum_{\mathbf{p} \in B_i} |\nabla_g(\mathbf{p})| \cdot \exp\Big(\frac{1}{m_i + n_i} \sum_{\mathbf{q} \in (B_i \cup E_i)} \mathcal{F}_p(\mathbf{q})\Big). \tag{2}$$

It is worth noting that an exponential function is used in Eqn. 2 to emphasize the significance of the pixels' focusness values. Since the above calculation does not apply directly to image margins, we manually assign fixed negative values to margin pixels by assuming low saliency.

After the focusness is computed for a region, we assign this value to every pixel in it. By doing this, we obtain a focusness map over the whole image I, which we denote as $\mathcal{F}(I)$ or $\mathcal{F}$ for short.

It is noteworthy that our region-level focusness computation essentially corresponds to a propagation of the focusness and/or sharpness at the boundary and interior edge pixels to the whole area of a region. Compared with the previous propagation methods [31, 41], ours is simple, stable and able to process regions with non-smooth interiors.

3.3. Objectness Estimation

Human eyes tend to identify an object as either salient or not as a whole. Therefore, it is desirable to estimate the probability of each region belonging to a well identifiable object in order to prioritize the regions in salient region detection.

Figure 3. From left to right: source images, uniqueness, focusness and ground truth.

Recently, Alexe *et al.* [3] proposed a novel trained method to compute an objectness score for any given image window, which measures the probability of that window containing a complete object. The objectness measure is based on image cues such as multi-scale saliency, color contrast, edge density and superpixel straddling.

According to [3], an object as shown in an image usually has the following general properties:

- it has a well-defined closed boundary in space,
- its appearance is different from its surroundings, and
- it is sometimes unique and stands out saliently.

These properties match well our perception of saliency in general. As such, utilizing this work, we propose a method to measure the objectness of each region, resulting in a complete objectness map over the image. This is done in two steps: pixel-level objectness estimation and region-level objectness estimation, as detailed below.

Pixel-level Objectness. In order to compute the objectness of each pixel (*i.e.*, the probability of there being a complete object in a local window centered on each pixel), we randomly sample N windows over the image, and assign each window $\mathbf{w}$ a probability score $P(\mathbf{w})$ to indicate its objectness calculated by [3]. Thereafter, we overlap all the set of all windows, denoted as $\mathbf{W}$, to obtain the pixel-level objectness $\mathcal{O}_p(\mathbf{x})$ for each pixel $\mathbf{x}$ by

$$\mathcal{O}_p(\mathbf{x}) = \sum_{\mathbf{w} \in \mathbf{W} \text{ and } \mathbf{x} \in \mathbf{w}} P(W_{\mathbf{x}}), \tag{3}$$

where $\mathbf{w}$ denotes any window in $\mathbf{W}$ that contains pixel $\mathbf{x}$. We set $N = 10000$ in our experiment. Similar pixel-level objectness was used in [33] for image thumbnailing.

Region-level Objectness. For every region Λ_i, we compute its region-level objectness $\mathcal{O}_r(\Lambda_i)$ as:

$$\mathcal{O}_r(\Lambda_i) = \frac{1}{|\Lambda_i|} \sum_{\mathbf{x} \in \Lambda_i} \mathcal{O}_p(\mathbf{x}). \tag{4}$$

After the objectness is computed for a region, we assign this value to every pixel in it. By doing this, we obtain an objectness map over the whole image I, which we denote as $\mathcal{O}(I)$ or $\mathcal{O}$ for short.

3.4. Uniqueness Estimation

Uniqueness, *i.e.*, the color contrast feature, has been effectively used for saliency detection in previous works [7, 27, 29]. They have used either pixel-level or region-level uniqueness, but none has used both simultaneously. By contrast, we combine them to capture both macro-level and micro-level features in the image. Furthermore, in preparation for the region-level uniqueness estimation, we segment the image using a different approach which leads to adaptive region sizes and better segmentation results.

The uniqueness $\mathcal{U}_r(\Lambda_i)$ for a region Λ_i is computed as:

$$\mathcal{U}_r(\Lambda_i) = \sum_{1 \le k \le N, k \ne i} |\Lambda_k| D(\Lambda_k, \Lambda_i), \tag{5}$$

where $D(\Lambda_k, \Lambda_i)$ is the color distance metric between regions Λ_k and Λ_i in the *Lab* color space [7].

For generating superpixels, we use the meanshift algorithm [5] that is different than the graph-based one [9] used in [7]. Our choice is based on the advantages of meanshift in boundary alignment, shape adaptivity and region size control, we use it instead for image segmentation. For a better object-size adaptivity, we set the minimum region area parameter to one tenth of the foreground area in the binarized pixel-level objectness map as computed using Eqn. 3. The other two parameters, spatial bandwidth and the feature bandwidth are empirically set to 20 and 15, respectively. This difference in image segmentation method significantly boosts the performance over [7], as will be demonstrated in Section 4.

Since our computations of $\mathcal{F}, \mathcal{O}$ and $\mathcal{U}_r$ are all on the region level, they can work together to locate the overall structure of salient objects. However, they may sometimes miss small local color details that appear salient to human vision as well. Therefore, we incorporate the pixel-level uniqueness into the computation in order to capture those small color details that may be locally salient. For each pixel $\mathbf{x}$, its uniqueness $\mathcal{U}_p(\mathbf{x})$ is computed as

$$\mathcal{U}_p(\mathbf{x}) = \sum_{\mathbf{x}' \in I \setminus \{\mathbf{x}\}} D(\mathbf{x}', \mathbf{x}), \tag{6}$$

where $D(\mathbf{x}', \mathbf{x})$ is the color distance metric between pixels $\mathbf{x}$ and $\mathbf{x}'$ in the *Lab* color space [7].

Finally, we define the overall uniqueness map $\mathcal{U}$ as:

$$\mathcal{U} = \mathcal{U}_r + \mathcal{U}_p. \tag{7}$$

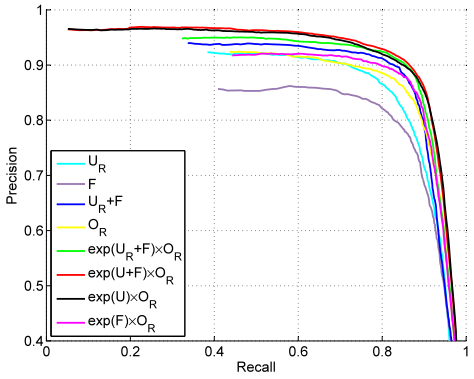


Figure 4. Evaluation of several combinations of our algorithm on the MSRA-1000 dataset [2].

3.5. Combination of Components

Combining the focusness, the objectness and the uniqueness maps as computed above, we define our final saliency map S as:

$$\mathcal{S} = \exp(\mathcal{F} + \mathcal{U}) \times \mathcal{O}, \tag{8}$$

where all operations are pixelwise, and $\mathcal{F}, \mathcal{U}$ and $\mathcal{O}$ are all normalized.

4. Experiments

4.1. Database and Evaluation Methods

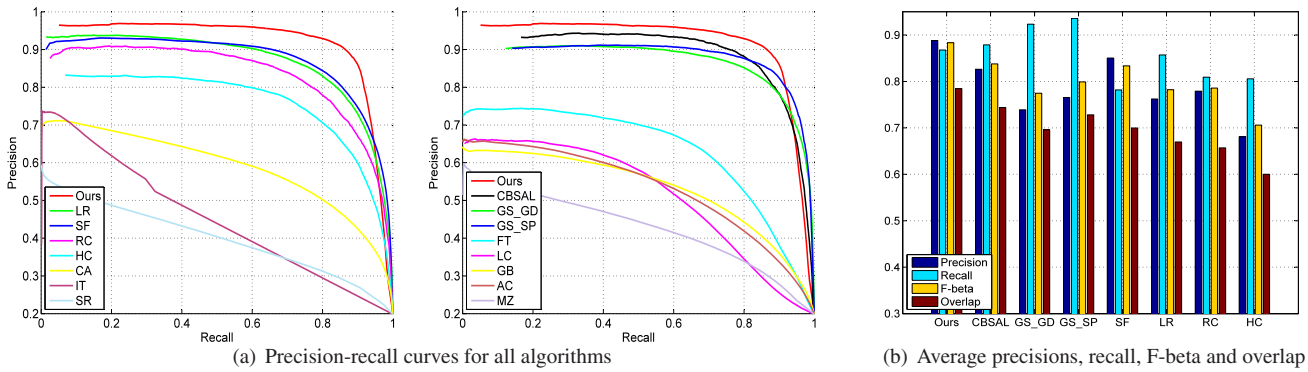
Our experiments are conducted on the MSRA-1000 dataset [2] and the BSD-300 dataset [25]. The MSRA-1000 dataset is a subset of the MSRA database [22], and each image in MSRA-1000 is associated with a human-labeled ground truth. The BSD-300 dataset contains many images with multiple (salient) objects and is considered the most difficult dataset in [4]. In order to study the performance of saliency detection algorithms, we use three popular evaluation protocols used in previous studies.

In the first protocol, we binarize each saliency map with a fixed threshold $t \in [0, 255]$. After the binarization, '1'-valued regions correspond to the foreground. We compare this binarized image and the ground truth mask to obtain the precision and recall. Varying t from 0 to 255 generates a sequence of precision-recall pairs, from which a precision-recall curve can be plotted. Combining the results from all the test images, we may obtain an average precision-recall curve.

In the second protocol, we compute the F-Measure as:

$$F_\beta = \frac{(1 + \beta^2) \times \text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}}. \tag{9}$$

We set $\beta^2 = 0.3$ as in [2, 7, 29, 34]. We also follow [2, 7, 29, 34] to use an adaptive binarization threshold t_a to binarize the saliency map before calculate the F_β. The threshold is



(a) Precision-recall curves for all algorithms (b) Average precisions, recall, F-beta and overlap

Figure 5. (a) Precision and recall rates for all algorithms on the MSRA-1000 dataset [2], results distributed into two sub-figures for better illustration. (b) Average precision, recall, F-beta and overlap using the adaptive thresholds on the MSRA-1000 dataset [2].

set as proportional to the mean saliency of the image:

$$t_a = \frac{k}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} \mathcal{S}(x, y), \tag{10}$$

in which we empirically choose $k = 1.5$. Furthermore, in order to comprehensively report the F-Measure characteristics, we uniformly sample a set of k in [0.1, 6] with an interval 0.1, compute the average F-Measure for each k, and then plot the average F_β-k curve correspondingly.

In the third protocol, we use the overlap rate to evaluate the saliency detection algorithms, which is defined as

$$R_o = \frac{F_g \cap G_t}{F_g \cup G_t}, \tag{11}$$

where F_g and G_t are the areas of the detected foreground and the marked ground truth, respectively.

4.2. Quantitative Comparison

On the MSRA-1000 dataset [2], we compare our method with other state-of-the-art approaches, including contrast-based approaches (IT [14], MZ [24], AC [1], FT [2], R-C [7], HC [7] and SF [29]), a graph-based approach (G-B [12]), a spectrum-based approach (SR [13]), priors-based approaches (GS [37], CBSAL [15]), and the ones with high-level priors (CA [11], LC [40], LR [34]). To evaluate these methods, we either use the results from the original authors (when available) or run our own implementations.

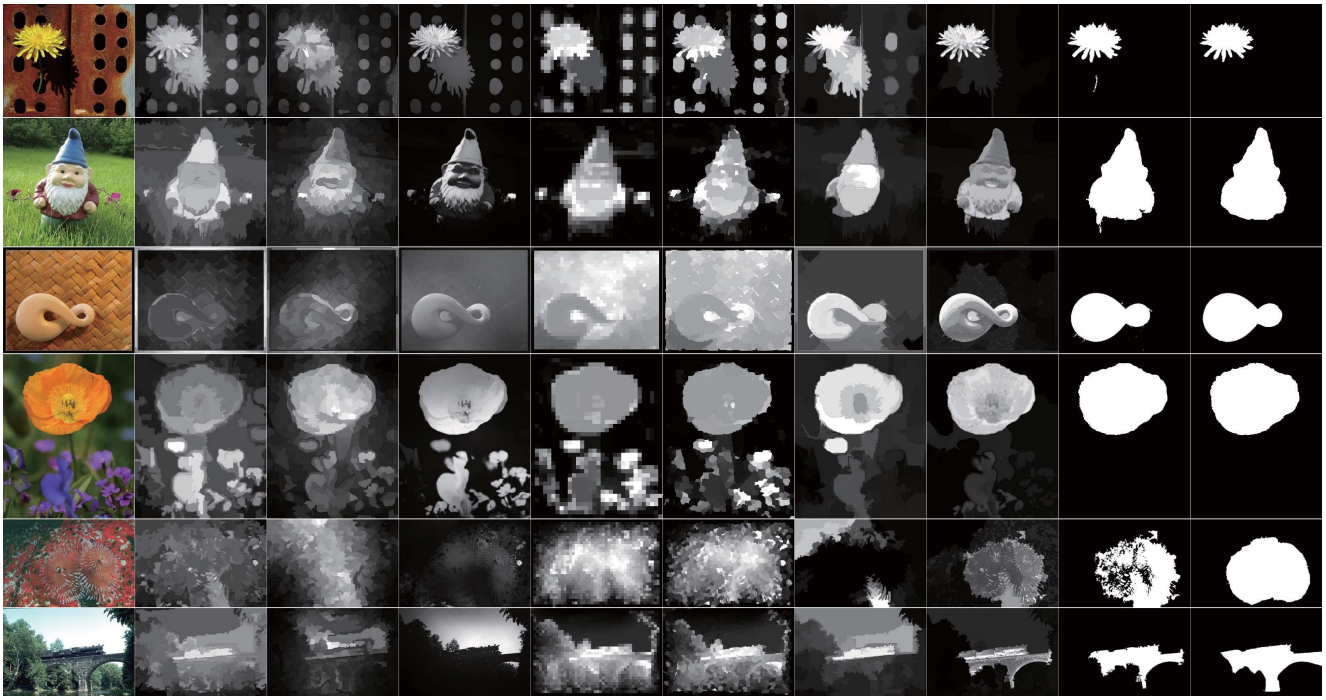
Fig. 5(a) shows the precision-recall curves of the above approaches on the MSRA-1000 dataset. As observed from Fig. 5(a), our method significantly promotes the precision and recall rate, and obtains the best result that maintains the recall rate at above 0.85 when the precision rate goes beyond 0.9, meaning that our method can get results close to the ground truth.

Besides, we compare the performance of various methods using the adaptive binarization threshold as computed in Eqn. 10 for each test image. As plotted in Fig. 5(b), we measure the average precision, the average recall, the average F_β and the average overlap for those methods. From Fig. 5(b), we see that, among the seven methods, ours has the highest precision, F_β and overlap and the third largest recall. This demonstrates the superior overall performance of our algorithm.

Furthermore, to comprehensively report the F-Measure characteristics, the F_β-k curves for various methods are shown in Fig. 7(a), from which we see that our method has the top F_β values at most selections of the k value.

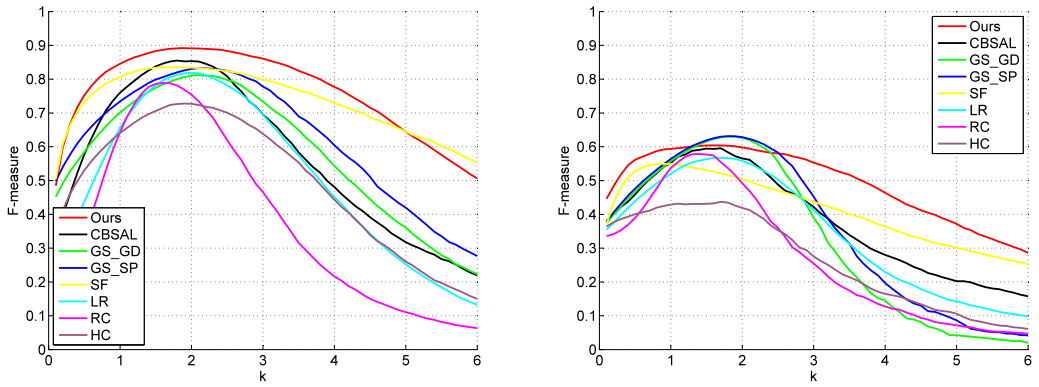
In order to demonstrate the effects of separate components and their combinations in our method, we plot their precision-recall curves in Fig. 4. From this figure, we see that region-level uniqueness ($\mathcal{U}_r$) and objectness ($\mathcal{O}$) both lead to better performance than focusness ($\mathcal{F}$), when applied alone. Nevertheless, the combination of region-level uniqueness and focusness ($\mathcal{U}_r + \mathcal{F}$) clearly boosts the performance of uniqueness ($\mathcal{U}_r$). After combining with the objectness ($e^{\mathcal{F}+\mathcal{U}_r} \times \mathcal{O}$), the performance is further improved. Finally, when the pixel-level uniqueness is incorporated ($e^{\mathcal{F}+\mathcal{U}} \times \mathcal{O}$), we obtain the top performance. We also illustrate the performance of ($e^{\mathcal{U}} \times \mathcal{O}$) and ($e^{\mathcal{F}} \times \mathcal{O}$) to demonstrate the importance of $\mathcal{F}$. It should be noted that $\mathcal{F}$ has already achieved better performance than HC [7] and $\mathcal{F} + \mathcal{U}_r$ has already achieved similar or better performance than the state-of-the-arts. Besides, uniqueness ($\mathcal{U}_r$) outperforms the RC method [7] in Fig. 5(a), though their only difference is in the image segmentation method, as discussed in Sec. 3.4. All of the above proves that the features we use are effective and complementary which lead to highly boosted performance when properly combined.

In order to evaluate the performance on detecting multi-object saliency, we also compare our method with most recent ones [7, 29, 34, 37] on the BSD-300 dataset [25]. The results are shown in Fig. 7(b) from which we observe outstanding performance of our method as well.



SRC RC [7] LR [34] SF [29] GS_GD [37] GS_SP [37] CBSAL [15] Ours BW GT

Figure 6. Visual comparison of previous approaches to our method and ground truth (GT). Due to space limit, only the results from six most recent other methods are presented. We also give the binary results of our method (BW) for adaptive threshold (Eq. 10). Our methods generate saliency maps closest to the ground truth.



(a) MSRA-1000 (b) BSD-300

Figure 7. (a): F_β-k curve on the MSRA-1000 dataset [2]. (b): F_β-k curve on the BSD-300 dataset [25].

4.3. Visual Comparison

Some examples for visual comparison of some approaches are shown in Fig. 6, from which we see that our method produces the best results on these images. Note that the test images of the top four rows and the last two rows are from MSRA-1000 [2] and BSD-300 [25], respectively.

We also compare the visual results from different components and their combinations in our approach. Some results are shown in Fig. 1, which confirm our intuition that it is all the components working together that leads to the best performance. Besides, the samples in Fig. 1 also show the failing mode of each component. In particular, uniqueness fails in the second row, objectness fails in the third row and focusness fails in the fourth row.

We give more samples in Fig. 3 to illustrate the failing mode of uniqueness and the capability of focusness.

5. Conclusions

In this work, we explicitly combine two important visual cues, focusness and objectness, with uniqueness for

saliency region detection. For the focusness estimation, we propose a novel method by scale-space analysis with solid mathematical proof; for the objectness estimation, we propose an effective window-overlapping-based approach utilizing one prior work on discrete window objectness estimation [3]. While uniqueness has been used by other works, we by contrast use both the pixel-level and the region-level uniqueness simultaneously to capture both macro-level and micro-level image features. More importantly, based on the complementary nature of the three visual cues, we combine them in an effective way that leads to the top performance when compared with the state-of-the-arts on two widely used public benchmark image datasets.

Acknowledgment

We thank reviewers for valuable suggestions to improve the paper. This work is supported by the National Natural Science Foundation of China (Grants No. 61070103 and No. U1035004), the Program for New Century Excellent Talents in University (NCET) in China, and Shandong Provincial Natural Science Foundation, China (Grant No. ZR2011FZ004). Ling and Yu are supported by US National Science Foundation (Grant IIS-1218156).

References

[1] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk. Salient region detection and segmentation. *In ICVS*, 2008. 2, 6

[2] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk. Frequency-tuned salient region detection. *In CVPR*, 1597–1604, 2009. 2, 5, 6, 7

[3] B. Alexe, T. Deselaers, and V. Ferrari. Measuring the objectness of image windows. *PAMI*, 2012. 1, 2, 4, 8

[4] A. Borji, D. N. Sihite, and L. Itti. Salient Object Detection: A Benchmark. *In ECCV*, 2012. 2, 5

[5] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *PAMI*, 24(5):603–619, 2002. 3, 5

[6] K. Y. Chang, T. L. Liu, H. T. Chen, and S. H. Lai. Fusing generic objectness and visual saliency for salient object detection. *In ICCV*, 2011. 2

[7] M. M. Cheng, G. X. Zhang, N. J. Mitra, X. Huang, and S. M. Hu. Global contrast based salient region detection. *In CVPR*, 409–416, 2011. 2, 5, 6, 7

[8] J. Elder and S. Zucker. Local scale control for edge detection and blur estimation. *PAMI*, 20(7):699–716, 1998. 3

[9] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient image segmentation. *In IJCV*, 59(2):167–181, 2004. 5

[10] C. Guo, Q. Ma, and L. Zhang. Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. *In CVPR*, 2008. 2

[11] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. *In CVPR*, 2010. 2, 6

[12] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. *In NIPS*, 545–552, 2006. 2, 6

[13] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. *In CVPR*, 2007. 2, 6

[14] L. Itti, C. Koch, and E. Niebur. model of saliency-based visual attention for rapid scene analysis. *PAMI*, 20(11):1254–1259, 1998. 2, 6

[15] H. Jiang, J. Wang, Z. Yuan, T. Liu and N. Zheng. Automatic salient object segmentation based on context and shape prior. *BMVC*, 2011. 2, 6, 7

[16] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li. Salient Object Detection: A Discriminative Regional Feature Integration Approach. *CVPR*, 2013. 2

[17] Z. Jiang and L. S. Davis. Submodular Salient Region Detection. *CVPR*, 2013. 2

[18] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In *ICCV*, 2009. 2

[19] C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Human neurobiology*, 4(4):219–227, 1985. 2

[20] T. Lindeberg. Edge detection and ridge detection with automatic scale selection. *In CVPR*, 1996. 3

[21] J. Li, M. D. Levine, X. An, and H. He. Saliency detection based on frequency and spatial domain analysis. *In BMVC*, 2011. 2

[22] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Y. Shum. Learning to detect a salient object. *PAMI*, 33(2):353–367, 2011. 2, 5

[23] C. Lang, T. Nguyen, H. Katti, K. Yadati, S. Yan, and M. Kankanhalli. Depth matters: Influence of depth cues on visual saliency. In *ECCV*, 2012. 2

[24] Y. F. Ma and H. J. Zhang. Contrast-based image attention analysis by using fuzzy growing. *In ACM Multimedia*, 374–381, 2003. 2, 6

[25] M. Movahedi and J. H. Elder Design and perceptual validation of performance measures for salient object segmentation. *In POCV*, 2010. 2, 5, 6, 7

[26] E. Niebur and C. Koch. Computational architectures for attention. *The attentive brain, Cambridge MA:MIT Press*, 163–186, 1995. 2

[27] Y. Niu, Y. Geng, X. Li, and F. Liu. Leveraging stereopsis for saliency analysis. In *CVPR*, 2012. 2, 5

[28] A. P. Pentland. A New Sense for Depth of Field. *In TPAMI*, 9(4):523–531, 1987. 3

[29] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung. Saliency filters: Contrast based filtering for salient region detection. In *CVPR*, 733–740, 2012. 1, 2, 5, 6, 7

[30] Y. Baveye, F. Urban, and C. Chamaret. Image and Video Saliency Models Improvement by Blur Identification. In *ICCVG*, 280–287, 2012. 2

[31] B. Soonmin and D. Fredo. Defocus magnification. *Computer Graphics Forum*, 26(7):571–579, 2007. 2, 3, 4

[32] J. Sun and H. Ling. Scale and object aware image retargeting for thumbnail browsing. *In ICCV*, 2011. 2

[33] J. Sun and H. Ling. Scale and Object Aware Image Thumbnailing In *IJCV*, 104:135–153, 2013. 4

[34] X. Shen and Y. Wu. A unified approach to salient object detection via low rank matrix recovery. In *CVPR*, 2012. 2, 5, 6, 7

[35] A. M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136, 1980. 2

[36] L. Wang, J. Xue, N. Zheng, and G. Hua. Automatic salient object extraction with contextual cue. *In ICCV*, 2011. 2

[37] Y. Wei, F. Wen, W. Zhu, and J. Sun. Geodesic saliency using background priors. In *ECCV*, 2012. 2, 6, 7

[38] Q. Yan, L. Xu, J. Shi, and J. Jia. Hierarchical Saliency Detection. In *CVPR*, 2013. 2

[39] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang. Saliency Detection via Graph-Based Manifold Ranking. In *CVPR*, 2013. 2

[40] Y. Zhai and M. Shah. Visual attention detection in video sequences using spatiotemporal cues. *ACM Multimedia*, 815–824, 2006. 2, 6

[41] S. Zhuo and T. Sim. Defocus map estimation from a single image. *Pattern Recognition*, 44(9):1852-1858, 2011. 2, 3, 4