Signal Processing: *Image Communication* 28 (2013) 1448–1463

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Superpixel based color contrast and color distribution driven salient object detection

Keren Fu [a,*], Chen Gong [a], Jie Yang [a], Yue Zhou [a], Irene Yu-Hua Gu [b]

[a] *Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China*

[b] *Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden*

ARTICLE INFO

Article history:
Received 7 March 2013
Received in revised form
17 July 2013
Accepted 18 July 2013
Available online 26 July 2013

Keywords:
Salient object detection
Saliency maps
Color contrast
Color distribution
Superpixels

ABSTRACT

Color is the most informative low-level feature and might convey tremendous saliency information of a given image. Unfortunately, color feature is seldom fully exploited in the previous saliency models. Motivated by the three basic disciplines of a salient object which are respectively center distribution prior, high color contrast to surroundings and compact color distribution, in this paper, we design a comprehensive salient object detection system which takes the advantages of color contrast together with color distribution and outputs high quality saliency maps. The overall procedure flow of our unified framework contains superpixel pre-segmentation, color contrast and color distribution computation, combination, and final refinement.

In color contrast saliency computation, we calculate center-surrounded color contrast and then employ the distribution prior in order to select correct color components. A global saliency smoothing procedure that is based on superpixel regions is introduced as well. This processing step preferably alleviates the saliency distortion problem, leading to the entire object being highlighted uniformly. Finally, a saliency refinement approach is adopted to eliminate artifacts and recover unconnected parts within the combined saliency maps.

In visual comparison, our method produces higher quality saliency maps which stress out the total object meanwhile suppress background clutter. Both qualitative and quantitative experiments show our approach outperforms 8 state-of-the-art methods, achieving the highest precision rate 96% (3% improvement from the current highest), when evaluated via one of the most popular data sets. Excellent content-aware image resizing also could be achieved using our saliency maps.

1. Introduction

Human usually pay more attention to some parts of a given image. This visual attention mechanism has been extensively studied by researchers, due to it can allow us to allocate our sensory and computational resources to the most valuable information. Salient object detection is one of the most important aspects of such attention mechanism. Various applications have been explored by using salient object detection, such as auto target location and segmentation [1,2], object based image retrieval [3], content-aware image resizing [4–7] and so on.

Saliency detection approaches usually can be categorized into two groups, so-called *bottom-up* and *top-down*. Bottom-up category [8,16–21] simulates our instinctive visual attention mechanism and lots of low-level features like color (intensity), edge (texture) could be adopted. Hence it is

* Corresponding author. Tel.: +86 138 170 83438.
E-mail addresses: fkrshichaoren@qq.com, fkrsuper@sjtu.edu.cn (K. Fu).

0923-5965/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.image.2013.07.005

stimulus and data driven. A salient object should be unique or have strong contrast compared to its surroundings with the respect of such features. Among them, color contrast is one low-level feature which may easily draw our attention [9,10].

The other category is called top-down [11,12,15]. Since top-down process in visual attention mechanism is defined as using effective *memory* to process presented *information*. Thus it is task and knowledge driven. Via the computer vision techniques, we can incorporate prior statistical knowledge or the high-level/object-level features such as faces, text, parts of human body detection or other task-specified object detection to simulate such kind of *memory*. As it has been proved that both bottom-up and top-down process contribute to human visual attention, some previous saliency models [13,14] belong to a mixture of the above two categories, i.e. combining both low-level and high-level features during detection.

Due to the adaptability of bottom up category, i.e. such kind of methods could be used to detect various target in more universal cases, in recent years, bottom-up methods have been widely yet deeply studied. Among lots of low-level features, color is the most informative one and might convey tremendous saliency information of a given image. Unfortunately, color feature is seldom fully exploited in the previous saliency models [16,19–21,34], in which only color contrast [16,19–21] or color distribution [34] is employed.

In this paper, we take the advantages of color contrast and color distribution to carry out our saliency detection, so our method mainly belongs to the *bottom-up* category, but slightly mixed with a *top-down color distribution prior*, which will be demonstrated in detail later. The previous work which is most related to ours should be [16] and recent [17]. The former defines pixel-wise saliency as a pixel's contrast to all other pixels. This is then converted into computation based on color histogram. Also, good results are reported using HC (Histogram Contrast) and RC (Region Contrast) methods in [16]. Unfortunately, their methods only consider the color contrast but exclude the color distribution, which may also be an important kind of character for salient object. So their HC and RC methods may not get good results on the images in which parts of background have relatively stronger contrast than the real salient object. Besides, the saliency maps obtained using RC usually pop out salient objects unevenly, and there is also lots of background clutter, as is shown in Fig. 4 (2nd and 3rd in the first row) and Fig. 8.

Our work also differs from recent [17] which measures the image saliency using element uniqueness and element distribution. In contrast, we measure the saliency based on the three disciplines for salient object in a more comprehensive way and we design a global saliency smoothing procedure which solves the ambiguity caused by distribution prior. Moreover, we refine the final saliency maps using meanshift segmentation to maintain the edge details, which facilitates the post-processing like object segmentation while [17] treats the refinement as a pixel-wise upsampling, aiming at improving the visual quality. As can be seen, our method outperforms [17] on providing more uniform saliency maps.

In summary, motivated by the previous work, we propose that a salient object may obtain the following characters on color feature at the same time, in two folds, the color contrast and color distribution.

I. The color components belong to a salient object may have strong contrast to their surroundings, which is biologically inspired [8]. (*contrast*)

II. These color components may be located near image center[1] rather than image boundary. It is based on the fact [13] that shows human fixation has much higher probability to fall onto the center area of the image. (*distribution*)

III. These color components usually distribute compactly [23]. In another word, color components which distribute widely are less likely to belong to a salient object. (*distribution*)

From the discussion, a salient object should obey the three disciplines above simultaneously. Thus the corresponding saliency could be formulated as

$$S = D^{prior} \cdot S^{contrast} \cdot S^{distribution} \tag{1}$$

in which S represents the final saliency of a pixel, a part or an object, while D^{prior}, $S^{contrast}$ and $S^{distribution}$ are respectively the corresponding distribution prior (**II**), color contrast (**I**) and color distribution (**III**) saliency. Since a salient object should obey the three disciplines above at the same time, a small value in any component (e.g. $S^{contrast} = 0$) may pull down the final saliency ($S = 0$).

Our method takes the above three characters to perform saliency detection. The experimental results show our algorithm can highlight a salient object more uniformly meanwhile keep a much cleaner background. As a result, high quality saliency maps can be obtained. Some examples are shown in Fig. 1. The contributions of this paper include:

1. A comprehensive salient object detection system is proposed, which combines the color contrast and color distribution in a unified manner.
2. Additional processing procedure such as global saliency smoothing and refinement using meanshift segmentation are introduced, which are proved to practically improve the performance of the system.
3. Our method outperforms the state-of-the-art in both qualitative and quantitative evaluations.
4. The results on other applications such as content-aware image resizing also show the superiority of our method against the recent approaches.

A preliminary conference version of this work appeared in [39].

The rest of this paper is organized as follows. Related works are described in Section 2. Our methodology is proposed in Section 3. Experimental results are analyzed in Section 4 while conclusion and future work are drawn in Section 5.

[1] Notice this character does not mean salient object should be just right fixed in the image center, yet means it usually appears in a centered area range, which could be represented using a probability-distribution-like prior map (Fig. 3).

Fig. 1. Saliency maps produced using our method.

2. Related work

As is mentioned above, the main categories of saliency detection methods are bottom-up and top-down. Since our method belongs to the former, here we only review related bottom-up category. For top-down category, we suggest readers to refer to a recent survey [35].

Among bottom-up category, as one of the earliest work, Itti et al. [8] proposed a center-surround operation as local feature contrast in the color, intensity, and orientation of an image. The center-surround operation is realized using DOG (Difference of Gaussians). Then Hou et al. [18] propose a method based on the spectral residual in the amplitude spectrum of Fourier transform. Zhai et al. [19] define the saliency of each pixel as its contrast to all other pixels. However, for efficiency, they only consider the luminance channel. Achanta et al. [20] propose a frequency tuned method which is extremely fast. They define the saliency of a pixel as its distance to the image average. But this algorithm is less promising for images that contain complex background and textures. Goferman et al. [21] combine local feature and global feature to estimate the patch saliency in multi-scale. This leads to high computational cost. Besides, the use of local feature may cause edges highlighted. Cheng et al. [16] propose Histogram Contrast based and Region Contrast based methods, called HC and RC respectively, as is mentioned in Section 1. Saliency maps obtained using their methods may contain background clutter and sometimes highlight parts of the object. Although they combine GrabCut [22] and their saliency maps to get good segmentation results, we demonstrate that high quality saliency map is the basis of various post processing. Thus the key point should be focused on how to improve the quality of the obtained saliency map. More recently, Perazzi et al. [17] combine color contrast and color distribution to perform saliency detection. They show that the complete contrast and saliency estimation can be formulated in a unified way using high dimensional Gaussian filters. Then, an upsampling procedure is carried out to assign each pixel a saliency value. Although better visual quality may be obtained, their saliency maps sometimes highlight only part of a salient object. Jung et al. [29] extent previous work [18] from global spectral residual into a local one and then combine these two into a unified spectral-domain approach. Although the local spectral residual provide a chance to analyze structural parts of a salient object, these parts are still corners or edges, leading to edges highlighted in the final saliency maps.

Furthermore, there are some bottom-up methods which adopt multiple features. Liu et al. [23] considers multiscale contrast, center-surround histogram, color spatial distribution as well as motion cue in a CRF learning manner. Gopalakrishnan et al. [30] models both the color and the orientation distribution in a given image and compute the two saliency maps separately. The final saliency map is selected as either color saliency map or orientation saliency map by automatically identifying which maps leads to correct detection results. Alexe et al. [24] propose the *objectness* and use a sliding window and compute a multiple low-level feature based saliency score for each window. Salient object corresponds to the window with the highest score. Feng et al. [25] compute the window saliency based on superpixels. They use all the superpixels outside the window to compose the inside ones, thus the global image context is combined. Fang et al. [31] present a method to perform saliency detection in compressed domain. They extract intensity, color, and texture features of the image from the discrete cosine transform (DCT) coefficients in the JPEG bit-stream. Lang et al. [32] render the saliency maps by finding the consistently sparse elements from the joint decomposition of multiple-feature matrices into pairs of low-rank and sparse matrices.

Our saliency detection method differentiates from the state-of-the-art bottom-up methods on expectation, for our method concentrates on how to produce high quality saliency maps which highlight the entire object uniformly as well as suppress the background clutter enormously. High quality saliency maps facilitate most post processing like object segmentation and content-aware image resizing, as we will show later.

3. Methodology

Fig. 2 shows the whole procedure flow of our method, including SLIC superpixel pre-segmentation [26], color contrast and color distribution computation, combination and final refinement. Because high quality saliency maps are produced, using simple thresholding may achieve good segmentation results, as we will show in the final quantitative comparison. Note that an input image is first resized to the size of (W, H), which subjects to $\max(W, H)=400$. W and H are respectively the width and height of the resized image. Then all the parameters of our method are tuned on this basic resolution.

3.1. Pre-segmentation

In order to calculate color contrast of a pixel to all other pixels, a straightforward way is pixel-wise computation, as is mentioned in [19]. However, the computational cost of such algorithms is $O(M^2)$, where M denotes the number of pixels in the input image. This is terrible because an input image usually contains hundreds of thousands pixels. An elegant way to speed up and reduce computational cost is histogram based computation [16,19] or segmenting images into edge-preserving regions, like that in [16,17,25]. Pixels in the same region usually have homogenous color component. Computing region based contrast instead of pixel-wise operation enormously pulls down the computational complexity. Thus, we first use SLIC superpixels [26] to decompose an image and generate spatial compact regions R_i, $i=1,2,3\ldots N$ with relatively consistent size, see Fig. 2 for instance. Compact SLIC superpixels are generated iteratively using mean-shift clustering based on the initial uniformly distributed region seeds. We use SLIC superpixels in LAB color space, as is suggested in [26]. For an input image, we segment it into about $N=500$ superpixels, a tradeoff between computational cost and description ability, and then use these superpixels as processing units. Here let c_i, p_i denote the average color and position of the ith superpixel

$$c_i = \frac{\sum_{I_m \in R_i} I_m^C}{|R_i|}, \quad p_i = \frac{\sum_{I_m \in R_i} I_m^P}{|R_i|} \tag{2}$$

where I_m^C and I_m^P are respectively 6D color vector, constituted by LAB and RGB components (corresponds to c_i^{LAB} and c_i^{RGB}), and position vector of pixel I_m. $|R_i|$ is the sum area of superpixel region R_i. Pre-segmenting the image into superpixels eliminates unnecessary details and noises as well.

3.2. Color contrast

According to character (**I**), we define the color contrast saliency of region R_i as

$$S_i^{contrast} = \sum_j D(c_i, c_j)^2 w_{ij}^P \tag{3}$$

in which $D(c_i, c_j) = \|c_i - c_j\|_2$ is Euclidean distance between c_i and c_j. That means we first combine both color systems in color contrast computation. This is motivated by one color system does not always work [27,33], since the LAB

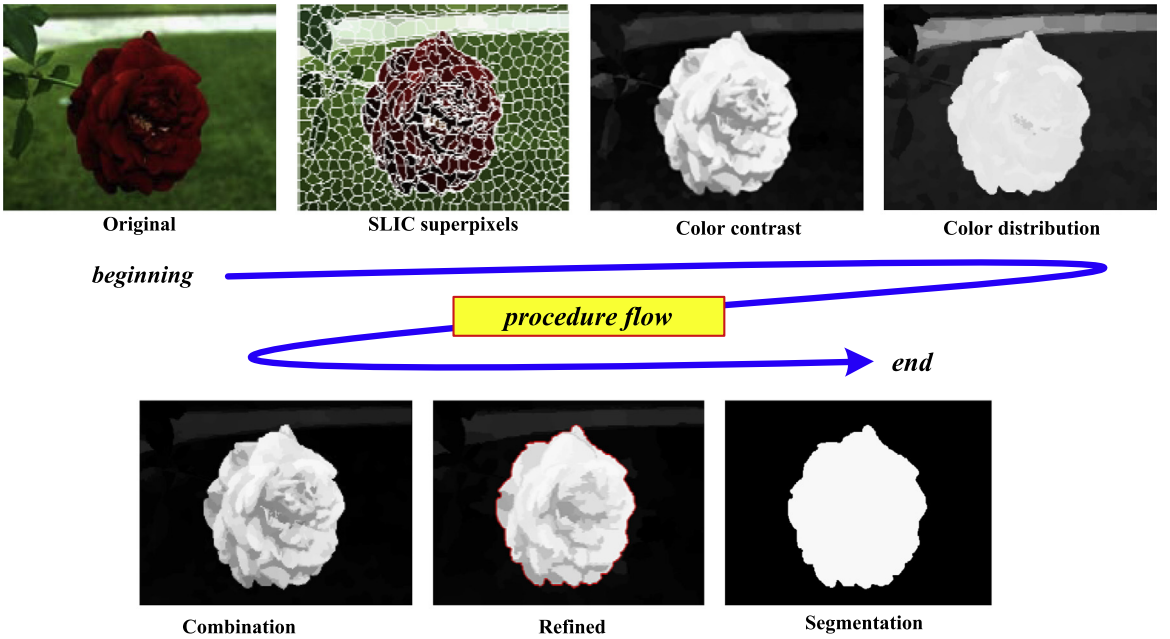

Fig. 2. The whole procedure flow of our method, including SLIC superpixel pre-segmentation, color contrast and color distribution computation, combination, final refinement and object segmentation. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

color space is well known on best characterizing the human visual perception while the RGB color system is the most widely used on variety of devices and displaying facilities. Different from [27] which computes saliency maps on different color systems and then averages the results, for convenience, we directly use the concatenated features vectors. Note that c_i and c_j is normalized (subtracting the mean and dividing the standard deviation) before computation in order to zoom different color system into the same scale. Here quadratic term of the 2-norm is used to better suppress the low contrast component (usually background). $w_{ij}^P = e^{-\alpha\|p_i - p_j\|^2}$ is spatial constrain, which enhances the effect of nearer neighbors. α controls the weight's sensitivity to spatial distance. When $\alpha \to 0$, $w_{ij}^P \to 1$, (3) degrades into global contrast calculation which is similar to that in [16]. Note the main difference between (3) and the region contrast (RC) proposed in [16] is that [16] uses segmented regions that are not enforced to be of roughly consistent size, thus the color contrast measure should involve region sum area. Since our superpixels are generated using SLIC which guarantees compactness and size consistency, thus it guarantees more accurate computation of color contrast and distribution.

Actually, only using color contrast may hardly filter out the false positives which have high color contrast but belong to the background clutter, e.g. high color contrast clutter near the image boundary (the white road in Fig. 4, contrast assumption fails to detect the flower but renders the road with the highest saliency in the 4th image). So in order to make our system more robust against such clutter, we introduce a *top down color distribution prior*, which meets character (**II**) mentioned in Section 1. An advantage of using distribution prior is that we could rule out the background color components which have the similar or even higher contrast than the color components that belong to the real salient object, since in practice, a photographer may seldom put the target near the image boundary while make the image center contain high contrast background. Although such prior may fail in abnormal or rare cases, it eliminates more false positive in normal cases. Notice that in Fig. 4, compared with color contrast based methods HC and RC, our approach renders the white road in the background much lower saliency. After combining the distribution prior, (3) should be rewritten as

$$\widehat{S}_i^{contrast} = D_i^{prior} \sum_j D(c_i, c_j)^2 w_{ij}^P \tag{4}$$

In (4), $\widehat{S}_i^{contrast}$ indicates an distribution-prior-enhanced version differentiating from (3). D_i^{prior} is the distribution prior of superpixel R_i. According to the fact that human fixation has much higher probability to fall onto the center area of the image, D_i^{prior} is larger when p_i is closer to the image center. Here a Gaussian distribution like $D_i^{prior} = e^{-w\|p_i - c\|^2}$ may be used, in which c denotes image center. However, instead of being puzzled by how to adjust parameter w, which controls the probability distribution of salient object's occurrence, a simple but effective statistical approach is used. We compute the average of 1000 ground truth images provided by [20] (Fig. 3). The ground truth images are first resized to resolution 400×400, and then are summed together and averaged. The average image commonly shows where a salient object is most likely to appear. Then it is normalized to have maximum value 1 to form the distribution prior map. In our implementation, the distribution prior map (resolution 400×400) is resized to the input resolution when it is used, and D_i^{prior} is directly obtained from the resized distribution

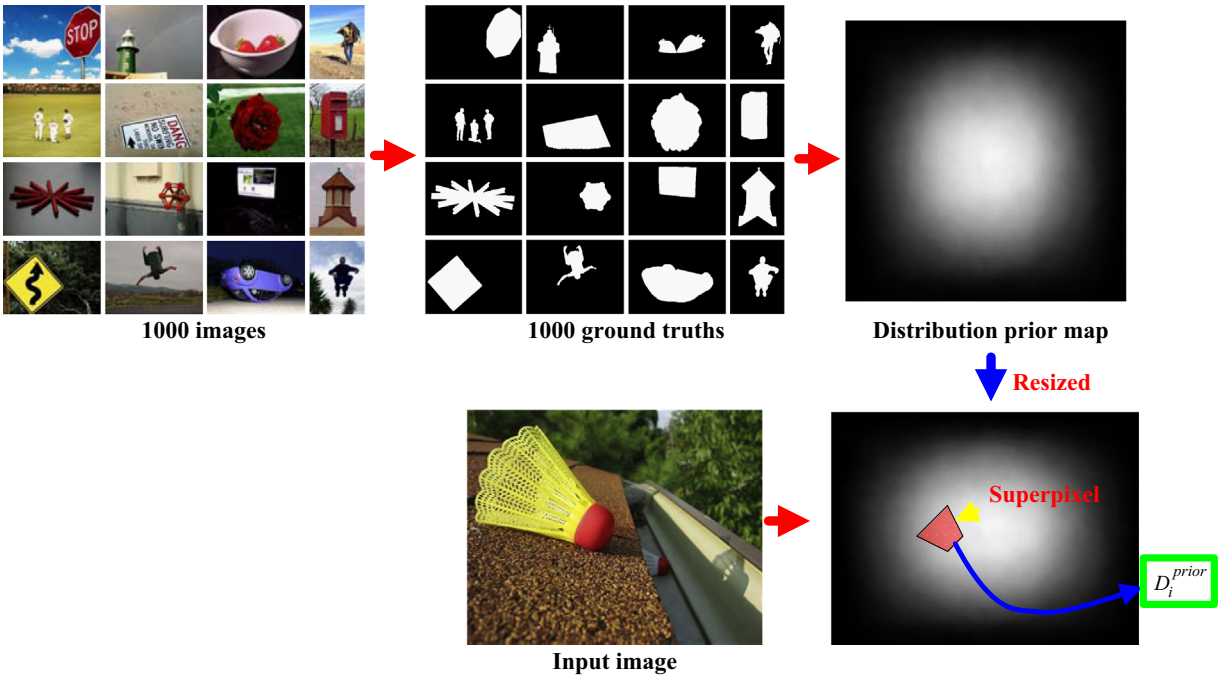

Fig. 3. An illustration for obtaining distribution prior map.

prior map (denoted as *RDM* for short) as

$$D_i^{prior} = \frac{\sum_{Im \in R_i} RDM(I_m^P)}{|R_i|} \quad (5)$$

where I_m^P is position vector of pixel I_m and $RDM(I_m^P)$ is the value on the resized distribution prior map at position I_m^P. Computing the average prior in a superpixel yields a more robust prior estimation. Actually when we compute the saliency map of an image from this dataset [20] (called MSRA-1000), we have considered the other 999 images' ground truth to calculate the distribution prior. This is reasonable since we could suppose images are usually independent from each other, hence independent from the computed prior map. We believe this is better than naively setting the parameter w in a Gaussian distribution to simulate such center bias. Besides, we find that the average of 999 images is almost equal to that of 1000 images, leading to nearly no influence on the final results. Distribution prior in Fig. 3 reveals salient objects usually appear in a centered area range, and we also find our prior map is compatible with the one-third rule in professional photography. We sample the four intersection points of one-third lines used in the professional photography and the resulting distribution prior values are respectively 0.6270, 0.6272, 0.6275 and 0.6275 whereas the points that are very close to the four boundaries (Fig. 3) acquire nearly zero prior values. This indicates the four intersection points obtain high prior values as well.

Combining color contrast and distribution prior may result in ambiguous/distorted saliency map and could not pop out the entire object uniformly, as is shown in Fig. 4 (the 5th image). This is caused by the fact that prior only highlights the central parts of the image. Even a uniform color contrast map being multiplied by the prior map would result in such phenomenon. As a solution, we propose to adopt a global saliency smoothing procedure in color space to assign closer saliency values to regions with similar colors as

$$\overline{S}_i^{contrast} = \sum_j w_{ij}^C \widehat{S}_j^{contrast} \quad (6)$$

where $\overline{S}_i^{contrast}$ represents the smoothed saliency. $w_{ij}^C = (1/N^C)e^{-\beta\|c_i^{LAB} - c_j^{LAB}\|}$ is the weight corresponding to color similarity, as LAB is better for smoothing in practice. $N^C = \sum_j e^{-\beta\|c_i^{LAB} - c_j^{LAB}\|}$ is its normalization term that guarantees all weights summed to 1. In our experiment, we find that exponent function works better than Gaussians on smoothing saliency of the whole object. In contrast, Gaussians fall down too sharply and usually highlight parts of object. β controls the extent of smoothing. When $\beta \to 0$, $w_{ij}^C \to 1/N^C$, after computing (6), all regions will obtain the same saliency, achieving the most extreme case. When $\beta \to \infty$, output $\overline{S}_i^{contrast}$ equals to $\widehat{S}_i^{contrast}$. Through computing (6), the whole object's saliency becomes more uniform (last image in Fig. 4). Note our global saliency smoothing is also different from the conventional saliency propagation mentioned in [34], since the proposed method in [34] uses Page Rank to propagate the saliency energy to obtain uniform saliency map. In contrast, we show this purpose could be achieved through a global computational smoothing fashion. By the way, an independent but related work is proposed in [38], which uses guided filter to smooth the region-based saliency map. However, since they have adopted segmentation method which is the same as that in [16], the generated regions with irregular size may degenerate the performance of the guided filter, which involves region location. For comparison, aiming at more uniform saliency rendering, our smoothing is global and we use SLIC superpixel to

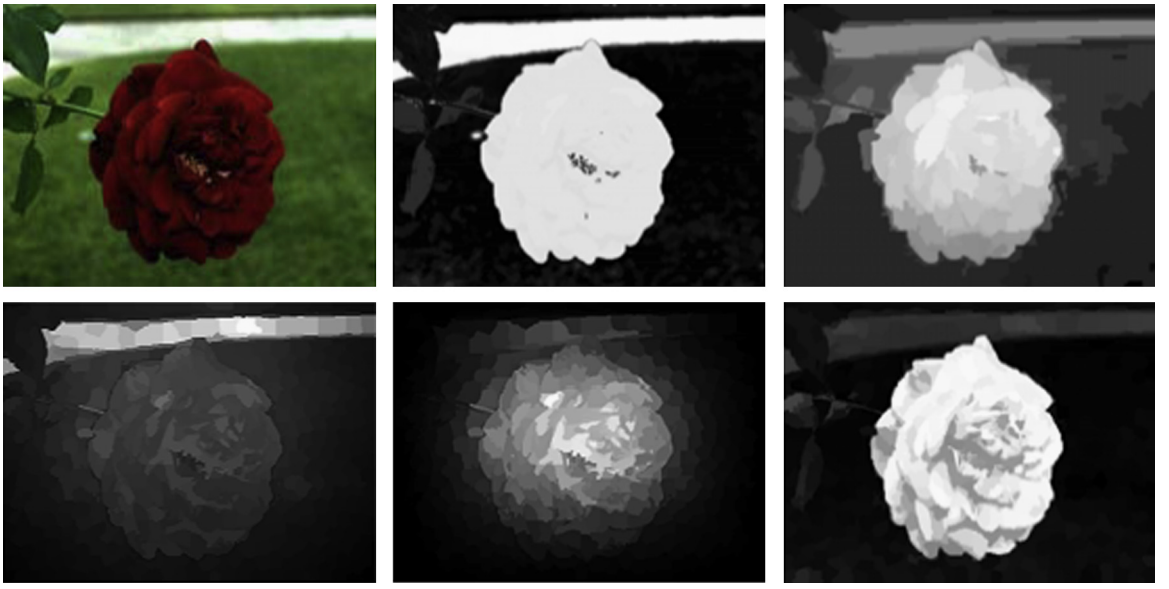

Fig. 4. From left to right, top to bottom: original image, result from HC, result from RC, contrast saliency without distribution prior, contrast saliency incorporating distribution prior, contrast saliency further incorporating global saliency smoothing. Note the distribution prior help filter out the high contrast clutter from the background while after the smoothing procedure, the overall object is highlighted uniformly. Notice that the road in the background is rendered the highest saliency by HC while RC highlights the flower unevenly.

pre-segment images, which guarantees compactness and size consistency as aforementioned.

The smoothed saliency map is then normalized to [0, 1] using linear stretch as (7) to get the ultimate color contrast saliency map.

$$\bar{S}_i^{contrast} \leftarrow \frac{\bar{S}_i^{contrast} - \min_j(\bar{S}_j^{contrast})}{\max_j(\bar{S}_j^{contrast}) - \min_j(\bar{S}_j^{contrast})} \tag{7}$$

In addition, we examine the smoothing power of β by choosing various β to obtain the smoothed color contrast saliency maps. Fig. 5 shows a typical example where β varies from 10 to 10^{-4}. When β get smaller, the performance becomes better. However when β is smaller enough, tuning it smaller again results in nearly no difference. This is because when β turns relatively smaller enough, the $e^{-\beta x}$ can approach linear tendency for a certain interval of x, that is $e^{-\beta x} \to 1-\beta x$ when $\beta x \to 0$. Thus a small number of β may satisfy the demand of uniform rendering. From another sight, the smoothing procedure also could be deemed as weighted averaging which helps to suppress background clutter. For example, suppose there are some clutter on the green grassland in Fig. 4 (means some parts of the grass are rendered high contrast saliency using aforementioned operation). However, after the saliency smoothing operation, the entire saliency energy of these parts will be assigned uniformly (i.e. is averaged) to the whole grassland. Thus each part will obtain very low saliency. This also could be concluded from Fig. 4 that before saliency smoothing procedure, although the saliency energy is concentrated on the flower, there are still some clutter from the grassland. However, after smoothing, the grassland is rendered the lowest saliency due to its largest area in the image. From the discussion above, the specially designed smoothing procedure may pop out the complete object meanwhile suppress background.

3.3. Color distribution

We compute color distribution similarly to [23] to meet character (**III**). However, the difference is that we model the color distribution of each superpixel separately, which seems more suitable in our superpixel based framework, while [23] models the whole image color using Gaussian Mixture Model (GMM) in a Bayesian treatment. Here, when we consider a specific superpixel's (e.g. R_i) color distribution variance, we first find superpixels $R_j, j = 1, 2, 3, \ldots, N$ that are similar with R_i on color appearance. This is measured using color similarity w_{ij}^C between c_j and c_i. The distribution variance of color component c_i is then defined as

$$D_i^{distribution} = \| \sum_j w_{ij}^C p_j^2 - \left(\sum_j w_{ij}^C p_j \right)^2 \|_1 \tag{8}$$

Here, the square of $p_j = (x_j, y_j)^T$ denotes the element square of vector p_j, that is $p_j^2 = (x_j^2, y_j^2)^T$. Actually (8) calculates the distribution variance in x and y direction and uses 1-norm to add them up. Note the parameter β in w_{ij}^C of (8) is tuned differently from that in (6), resulting in more promising performance in practice. Eventually, $D_i^{distribution}$ is normalized to [0, 1]

$$D_i^{distribution} \leftarrow \frac{D_i^{distribution} - \min_j(D_j^{distribution})}{\max_j(D_j^{distribution}) - \min_j(D_j^{distribution})} \tag{9}$$

As demonstrated above, high distribution variance indicates that the corresponding color components are widely distributed in the whole image and are less likely to belong to a salient object, while low variance indicates a spatially compact distribution. Thus regions with high distribution variances should obtain low saliency, so we define the color distribution based saliency as

$$S_i^{distribution} = 1 - D_i^{distribution} \tag{10}$$

3.4. Combination and refinement

As we illustrated in (1), it is reasonable to non-linearly integrate the color distribution saliency and the color contrast saliency, as is presented in (11). Such non-linear combination can better pop out salient objects meanwhile suppress background than linear combination, just see Fig. 6 for examples.

$$S_i = \bar{S}_i^{contrast} \cdot S_i^{distribution} \tag{11}$$

After combination, there may still be noises and artifacts due to quantization errors of superpixel segmentation (see Figs. 6 and 7). In order to get high quality saliency maps, we again segment the images into spatially non-compact regions $R_k', k = 1, 2, 3 \ldots N'$ using mean-shift segmentation [28]. We set conservative parameters $sigmaS = 7, sigmaR = 6.5, minRegion = 240$ for all images to avoid under-segmentation. Then the image saliency is refined based on these regions as

$$S_k' = \frac{\sum_{I_m \in R_k'} I_m^S}{|R_k'|} \quad \text{s.t. } I_m^S = S_i|_{I_m \in R_i} \tag{12}$$

where I_m^S is pixel saliency computed by (11) (means if pixel I_m belongs to R_i, then its saliency I_m^S is equal to the

Fig. 5. The smoothing power of β. When β gets smaller, the smoothing is more powerful and the saliency of the whole object becomes more uniform. However, when β is smaller enough, the performance remains the same.

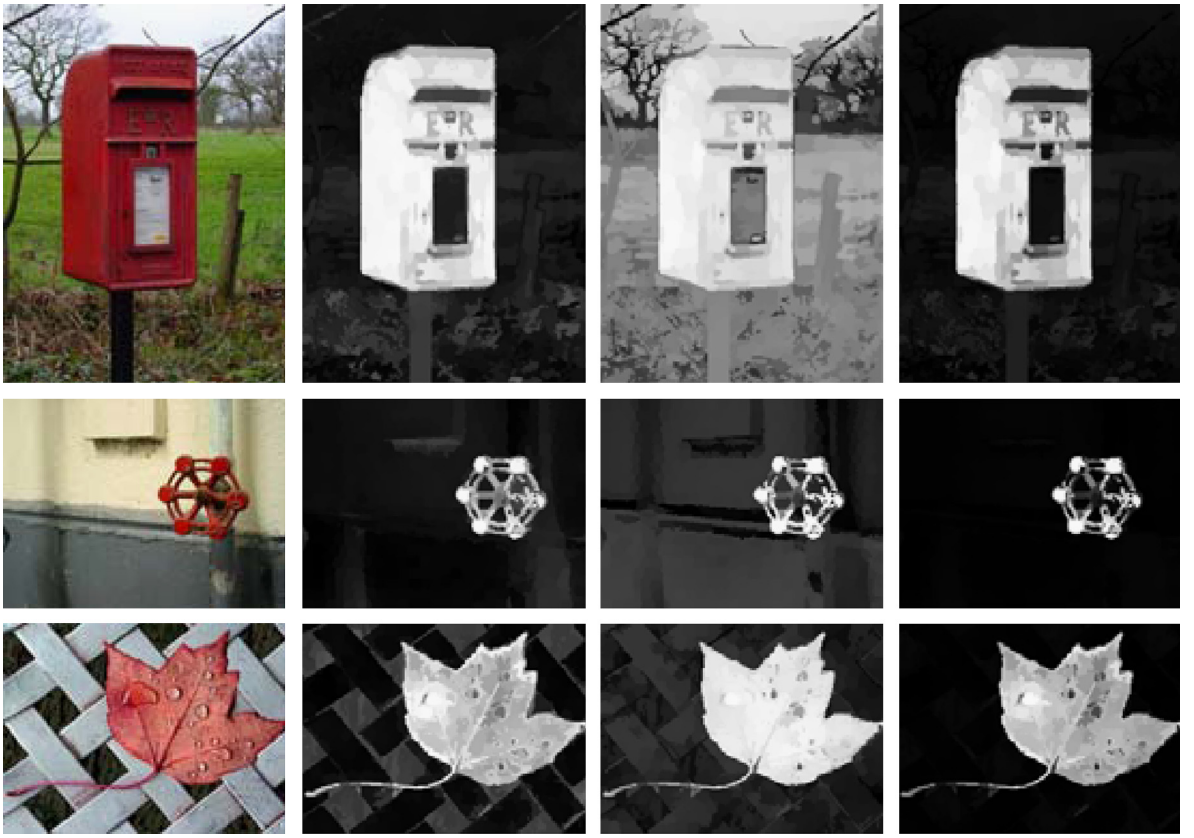

Fig. 6. Examples of combination. From left to right: original images, color contrast saliency maps, color distribution saliency maps, combined saliency maps via non-linear integration. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

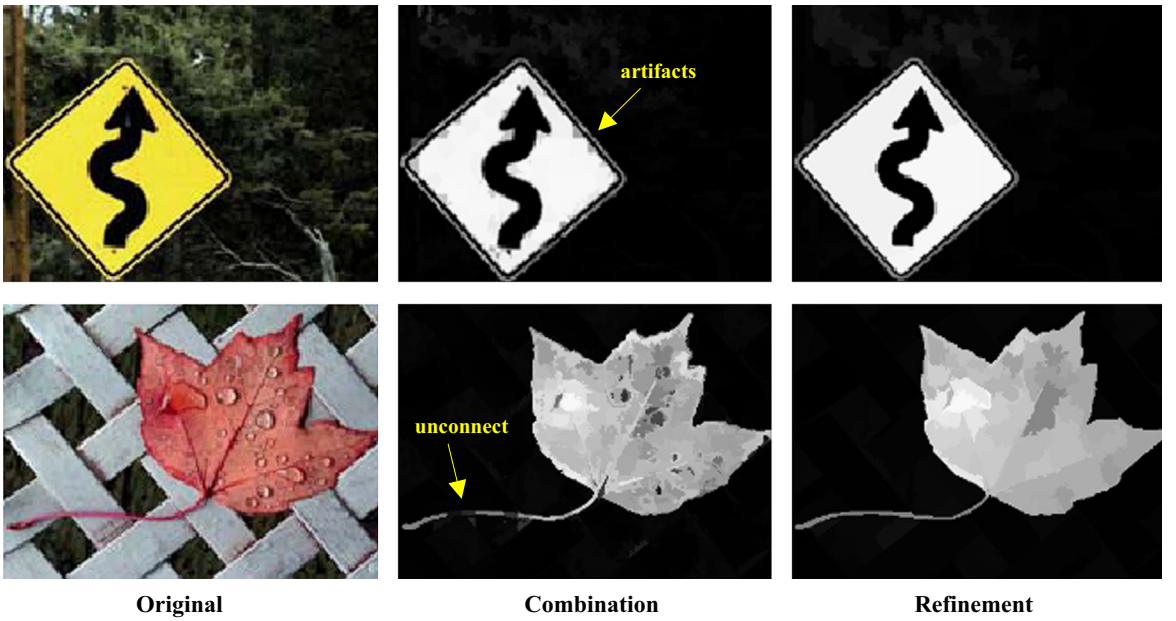

Fig. 7. Thanks to the refinement operation, artifacts are eliminated while unconnected parts generated by segmentation errors become connected.

superpixel's saliency S_i). $|R_k^{'}|$ is the sum area of region $R_k^{'}$. Compared to aforementioned pre-segmentation, mean-shift segmentation [28] may segment some homogenous object surfaces or background into large regions, hence further eliminating saliency inconsistency. In addition, since mean-shift segmentation is not enforced to produce

consistent size region, it adheres image boundary better, especially for thin elongated part of the object (petiole of the leaf in Fig. 7). Thanks to this refinement operation, artifacts are eliminated while unconnected parts generated by pre-segmentation become connected (Fig. 7). Finally, $S_k^{'}, k=1,2,3\ldots N'$ is normalized to [0, 1] to render our final saliency map.

4. Experiment and comparison

4.1. Parameter setup

Here, we summarize and review the crucial parameters of our system. Note all parameters below are tuned on the basic resolution which satisfies max $(W, H)=400$ and these parameters are kept consistent during the whole experiment. First the spatial constrain α in (4) should not be too large, because large value only takes nearby superpixels into consideration and merely highlights edge superpixels of large-scale object. Thus we conservatively set this parameter to a small value, e.g. 10^{-5}. We find the performance of our method does not change too much when using various such small values. The β in (6) is set to a relative small number (e.g. 10^{-3}). Tuning it smaller again has little impact on the detection results, as we show in Fig. 5. A relatively complicate parameter is the β in (8), which should not be too large or too small. Actually we tune this parameter (set to 10^{-1}) on a very small set (about tens of images) and find it generalize well for other images. In summary, our system only has three crucial parameters that need to be determined manually and the above values are also deemed as default in our system.

4.2. Evaluation on MSRA-1000 dataset

We test our method on the public dataset provided by [20]. This dataset is derived from the original MSRA dataset (5000 images) [23] and contains 1000 images with usually one unambiguous salient object in each image. The difference is that the dataset from [20] contains object-contour based ground truth while MSRA dataset [23] only provide bounding boxes as ground truth. Since object-contour based ground truth help conduct more accurate evaluations, to our best, this dataset is the most widely used for comparison between different saliency detection methods. We select the current popular 8 state-of-the-art saliency detection methods including IT [8], SR [18], CA [21], FT [20], LC [19], HC [16], RC [16] and SF [17] for comparison. The saliency maps of previous works excluding SF are provided by [16].[2] The SF [17] saliency maps are obtained from the author's webpage.[3] Fig. 8 shows several comparison results. Visually, it can be seen that our method obtains relatively higher quality saliency maps compared with the rest 8 methods, which sometimes highlight parts (RC and SF), corners (IT) or edges (SR and CA) with relatively more background clutter (HC, RC, LC and FT). In contrast, our method performs better on stressing out the complete prominent object while suppressing background.

Besides visual comparison, we also implement quantitative comparison. We evaluate the performance of our method by comparing its *precision–recall* rate. For a given threshold T, the *precision* and *recall* rate of a certain saliency detection method are defined as

$$Precision(T)=\frac{1}{1000}\sum_{i=1}^{1000}\frac{|M_i(T)\cap G_i|}{|M_i(T)|} \tag{13}$$

$$Recall(T)=\frac{1}{1000}\sum_{i=1}^{1000}\frac{|M_i(T)\cap G_i|}{|G_i|} \tag{14}$$

where $M_i(T)$ is the binary mask obtained by directly thresholding the saliency map using threshold T on the ith image. G_i is the ground truth. $|\cdot|$ denotes mask's sum area. As we use data set provided by [20], (13) and (14) are the averages of 1000 terms. In order to draw the precision and recall curves under different T, we use every possible threshold T from 0 to 255. This is similar to the fixed threshold experiment in [16,17,20].

The left sub-figure in Fig. 9 show the precision and recall curves. As can be seen, our method presents the best precision and recall curve. Our maximum precision rate is 96%, with 3% improvement from the second best 93% (SF). Another interesting phenomenon is that our method maintains high precision rate under various recall rate, i.e. our model is the best for recall smaller than 0.93 and is slightly worse than SF for recall between 0.93 and 1. This is actually consistent with our visual evaluation, which shows our approach provides high quality saliency maps that highlight the whole objects uniformly while suppress background.

As in many applications, high precision and high recall are both desired. Thus in addition to precision–recall curves, we also evaluate the *F-measure*, which is an integrated evaluation criterion that combines precision and recall as

$$F_\beta(T)=\frac{(1+\beta^2)Precision(T)\times Recall(T)}{\beta^2\times Precision(T)+Recall(T)} \tag{15}$$

where β^2 is set to 0.3, as is suggested in [16,17,20]. The right sub-figure in Fig. 9 show F-measure curves varying with threshold T. Compare with other methods, our method achieves high F-measure scores in a wide range, indicating less sensitivity to the threshold value. Notice that under this F-measure criterion, RC sometimes performs less better than HC. This may be attributed to that RC achieves higher precision, but lower recall at the same time, which pulls down the F-measure score to some extent. The same thing also happens on SF. As high quality saliency maps can be obtained using our method, using simple fixed threshold can achieve good segmentation results.

Besides, an adaptive threshold experiment, similar to that in [17,20], is carried out. The adaptive threshold T_a is defined as two times the mean saliency of an obtained saliency map, as is shown in (16).

$$T_a=\min\left\{2\times\frac{\sum_i^M S(I_i)}{M}, T_{max}\right\} \tag{16}$$

[2] http://cg.cs.tsinghua.edu.cn/people/~cmm/Saliency/Index.htm.
[3] http://graphics.ethz.ch/~perazzif/saliency_filters/.

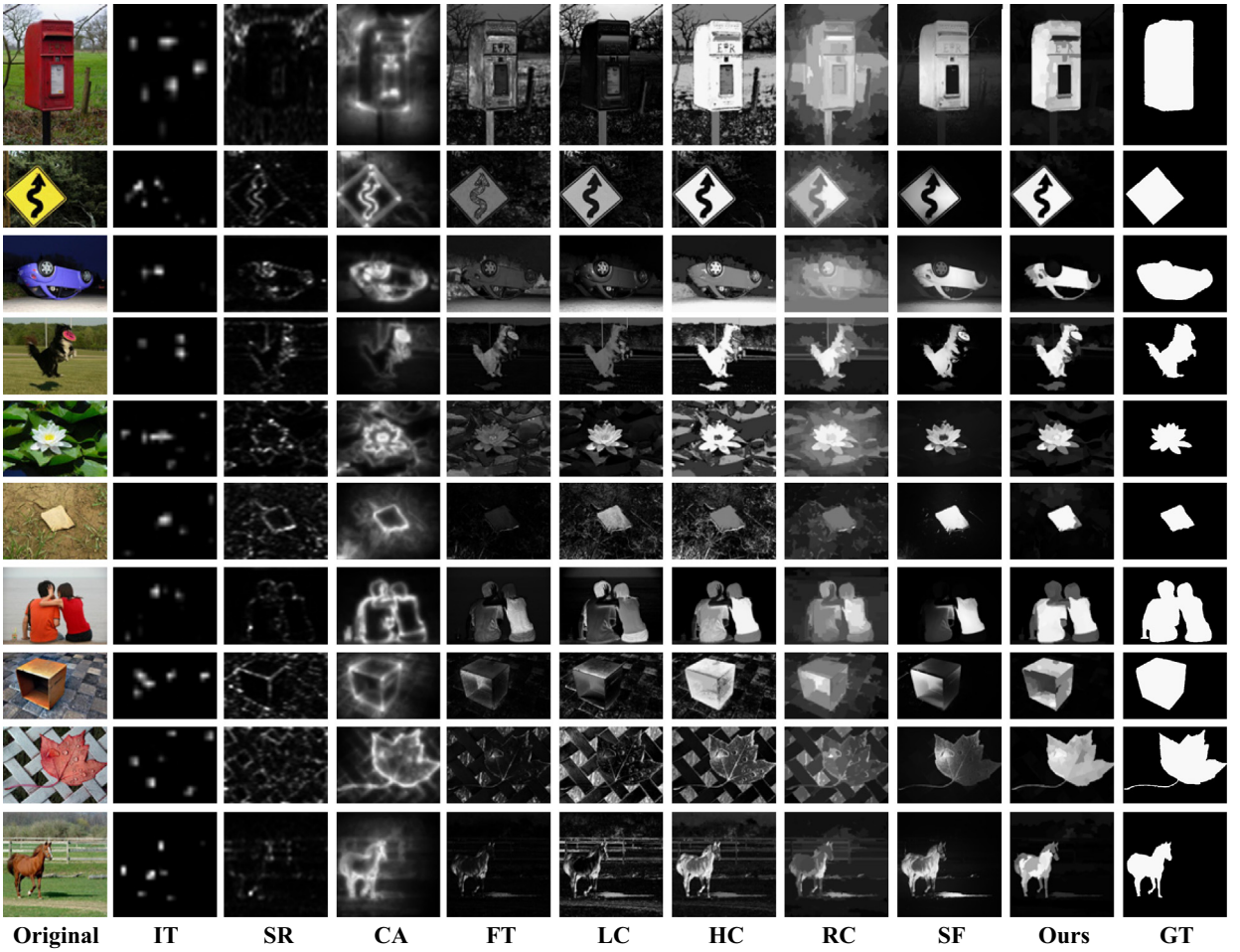

Fig. 8. Visual comparison results between our method and other 8 popular state-of-the-art methods.

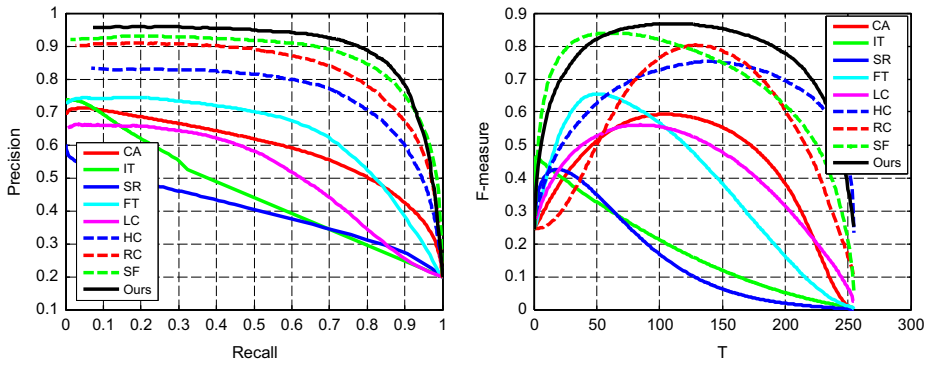

Fig. 9. Precision–recall, F-measure curves of 8 state-of-the art methods including CA, IT, SR, FT, LC, HC, RC, SF as well as our method on MSRA-1000 dataset.

where M denotes the number of pixels in the saliency map and i is pixel index. T_{max} is the upper bound for T_a and is set to 255 by us. Fig. 10 shows the precision, recall and F-measure in adaptive threshold experiment. It can be seen that in this experiment, RC still achieves high precision but low recall, because RC usually highlights only part of the real salient object. The precision rate of SF is very close to our method, but our method shows the highest recall rate and F-measure score, respectively 81% (9% improvement) and 0.86 (0.03 improvement).

4.3. Evaluation on complex SOD dataset

To evaluate the effectiveness of our method, we have done comparisons on SOD dataset [37] which contains 300 images with more complex background and also provides

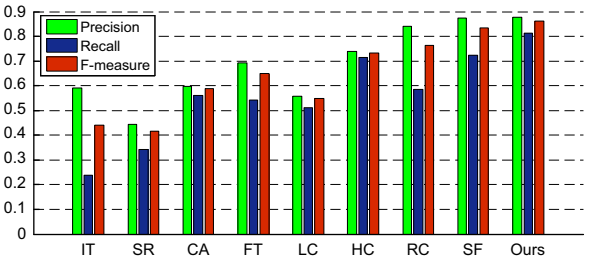

Fig. 10. Evaluation for adaptive threshold experiment.

foreground mask as groundtruth. Note here the prior map used is still the mask average map from MSRA-1000 dataset as shown in Fig. 3, since we just need a center bias hypothesis and there is no need to compute it specially for each testing dataset. We could also use this prior map for any other images from any other datasets. Fig. 11 shows the quantitative evaluations.[4] The accuracy of all methods are much lower, indicating that this database is much more difficult. Some simple contrast-based methods like LC, HC, FT could not handle such complex scenes and achieve relatively poor performance. The IT model which mainly uses Difference of Gaussians (DOG) and tends to highlight edges performs better than before since some objects from SOD contains complex texture and could be detected by DOG. The advantage of our method is that it could uniformly pop out the salient object meanwhile suppress background clutter, which could be observed from the visual comparison in Fig. 12. This may be benefited from taking the three characteristics of salient object into consideration at the same time and further combining saliency smoothing guarantees uniform saliency map. Thus our method achieves both higher precision and F-measure.

4.4. Evaluation for each individual step

Fig. 13 presents the evaluation for individual phase of our algorithm on MSRA-1000 dataset, respectively including only using color contrast saliency maps, only using distribution saliency maps, saliency maps without distribution prior, saliency maps without saliency smoothing and saliency maps without refinement. It shows the benefit of combining all steps while adding distribution prior and saliency smoothing really works for enhancing the performance on obtaining higher precision and wider range of high F-measure. Fig. 14 shows some intermediate results visually. It could be concluded form the 2th, 3th and 7th column that color contrast and color distribution generate complementary performance and the combination of both would achieve better results. 4th and 7th column show that employing the prior usually does not change the saliency detection results of some salient objects that are placed off the image center, yet it has a chance to correct the false positive near the image boundary (see 1st, 6th and 8th row). Again observing the 5th and 7th column, we could see that without the saliency smoothing procedure, the salient objects could not be popped out uniformly. 6th and 7th column conclude that refinement could offer fine boundary details while maintain saliency consistency.

[4] Because SF [17] only provides detection results on MSRA-1000 dataset, we could not compare with it on SOD dataset.

4.5. Boosting for the state-of-the-art and comparisons

Additionally we have explored the effectiveness of the distribution prior and refinement procedure exploited in this paper. We have conducted a boosting experiment similar to [36], in which we multiply the saliency maps of the above 8 methods with the distribution prior map in Fig. 3. Then the mean-shift refinement mentioned in (12) with the same parameters is used to further enhance the results. The evaluations on MSRA-1000 dataset are shown in Fig. 15 and some interesting conclusions could be drawn. First all 8 methods have a boost on performance compared to that in Fig. 9, which shows the effectiveness of the prior and refinement. Some methods like FT and HC even have more dramatic improvement than RC. The weak boosting of RC is caused by the most enormous clutter in its saliency maps, which could be inferred from Figs. 8 and 12. Such clutter brings two drawbacks into boosting. The one is the clutter in the image center gains unwanted enhancement by the prior. The other is the clutter of other uninteresting parts could not be suppressed efficiently by the prior (somewhat shown in Fig. 16). Although our distribution prior provides good accuracy, leading to remarkable improvement on some methods, these two drawbacks make RC have relatively lower precision under higher recall. This fact also reveals naively adding the center boosting cannot always boost the performance in a large gap. For comparison, our method still stands on the top and achieves high F-measure under wide range of T. Please note one thing that although after boosting using prior and refinement, the quantitative evaluation could obtain high scores, the visual performance may be degenerated, since such naive combination fail to highlight the object which is placed off the image center uniformly. Some visual results for this boosting experiment are shown in Fig. 16.

4.6. Content-aware image resizing

In content-aware image resizing [4–7], saliency maps are usually used to specify relative importance across image parts. These important content should be preserved to the original while the other unimportant pixels have to take more sacrifice. Here we use the framework proposed in [4] to validate the performance of the saliency maps produced by our method on smart image resizing task. For an image with resolution $[X, Y]$, we only consider the scaling along x-axis while other cases are straightforward. The resized resolution is denoted as $[X', Y']$, where $Y = Y'$ and $X' < X$. $\lambda = X'/X$ is called scaling factor. The λ in Fig. 17 is set to 0.5. Since our method generates uniform saliency maps compared to RC and SF, when used in content-aware resizing, our saliency maps better preserve the whole objects during scaling.

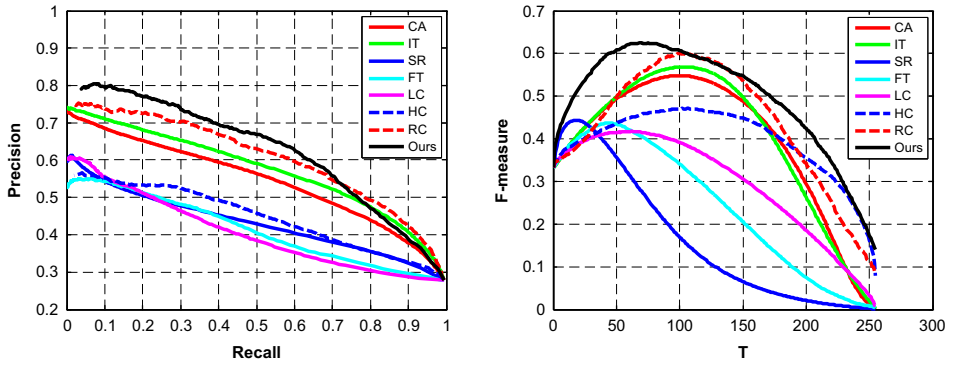

Fig. 11. Precision–recall, F-measure curves of 7 state-of-the art methods including CA, IT, SR, FT, LC, HC, RC as well as our method on SOD dataset containing images with more complex background.

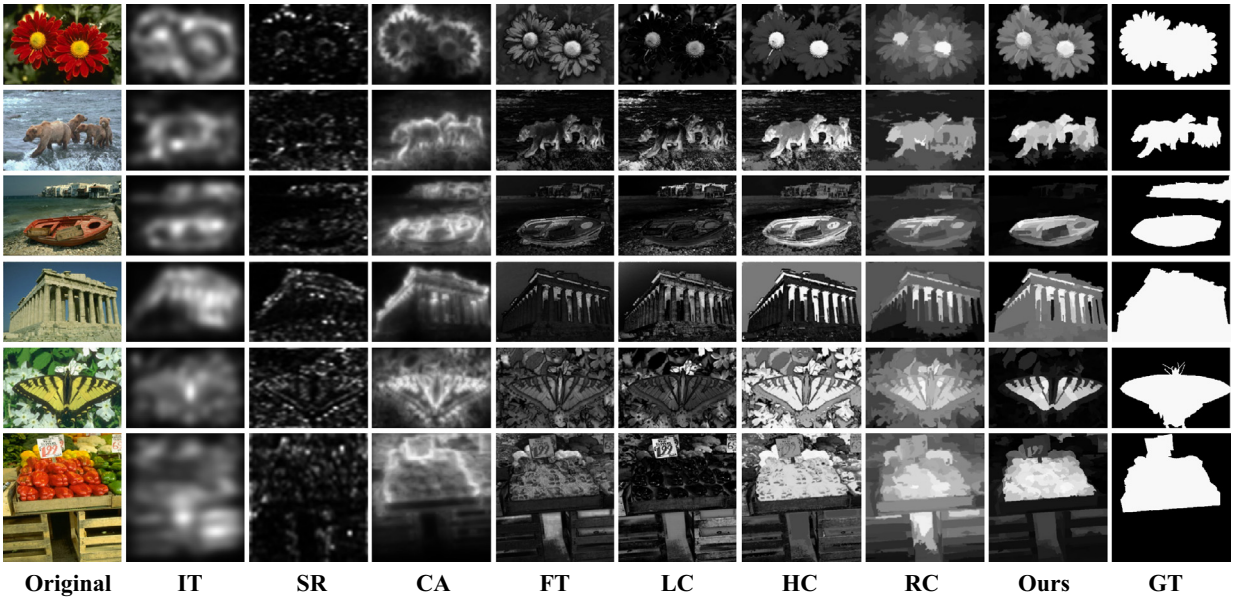

Fig. 12. Visual comparisons on SOD dataset.

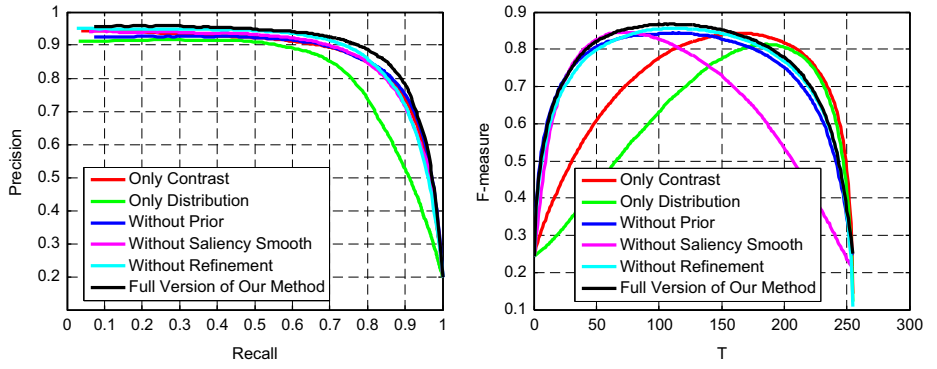

Fig. 13. Individual phase of our algorithm, respectively including only contrast, only distribution, without distribution prior, without saliency smoothing and without refinement.

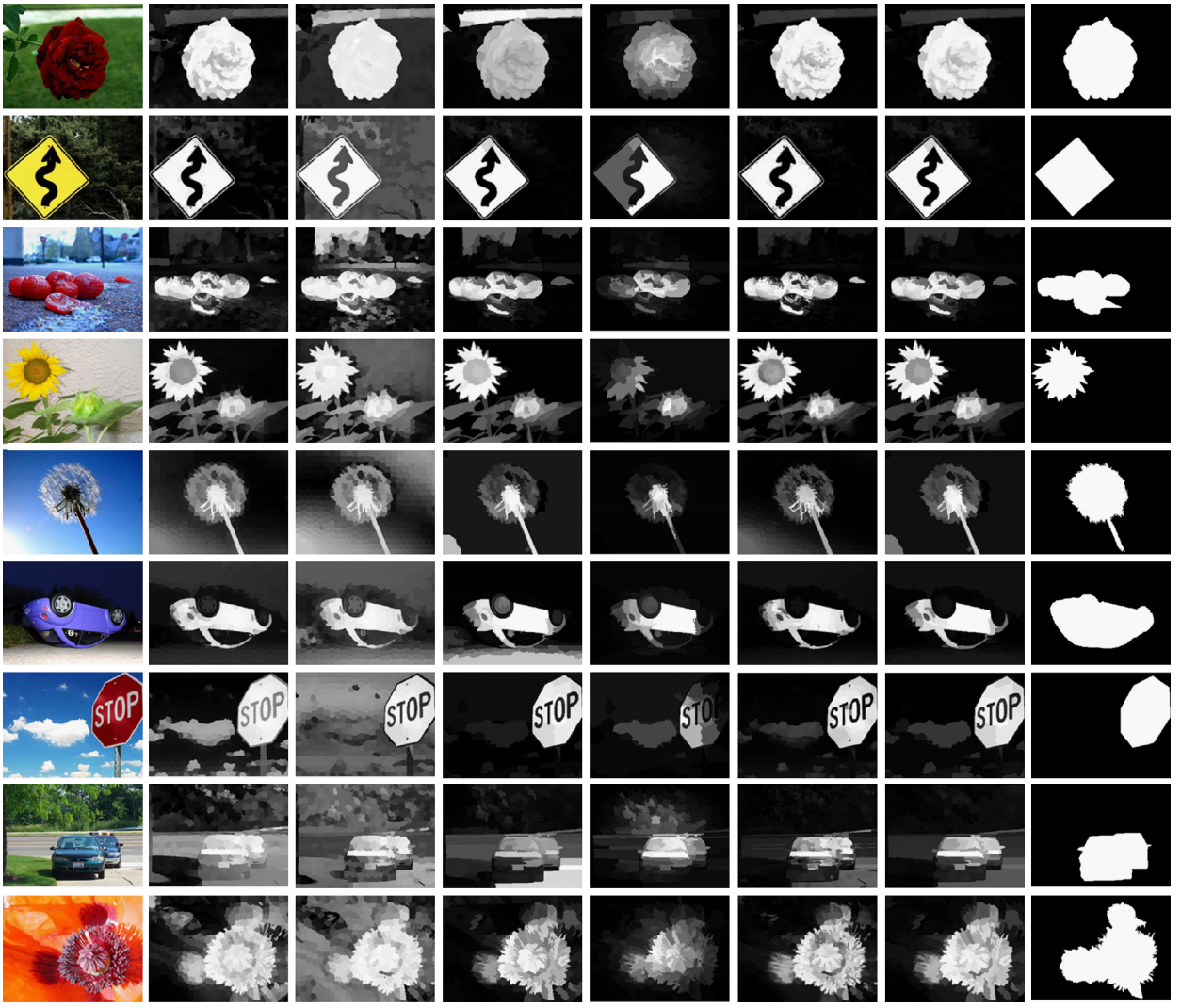

Fig. 14. Visualization of intermediate results. From left to right are respectively: original images, only color contrast, only color distribution, without prior, without smoothing, without refinement, full version of our method and ground truth. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

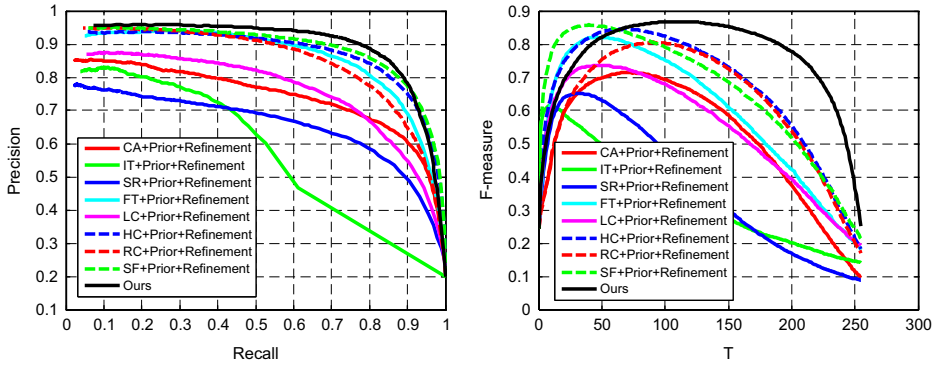

Fig. 15. Combining 8 methods with the prior and refinement procedure employed in this paper. All 8 methods have a boost on performance.

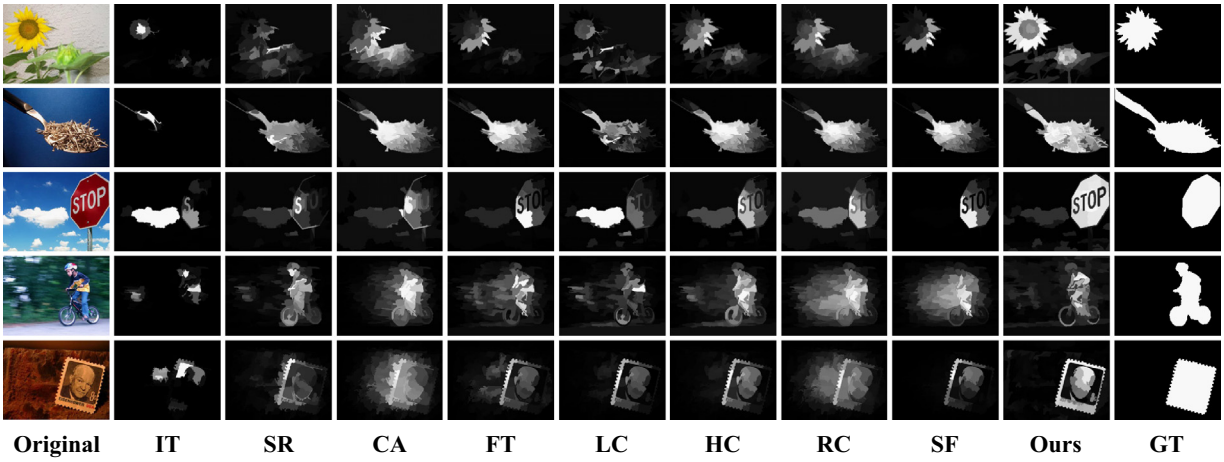

Fig. 16. Visual results for boosting experiment.

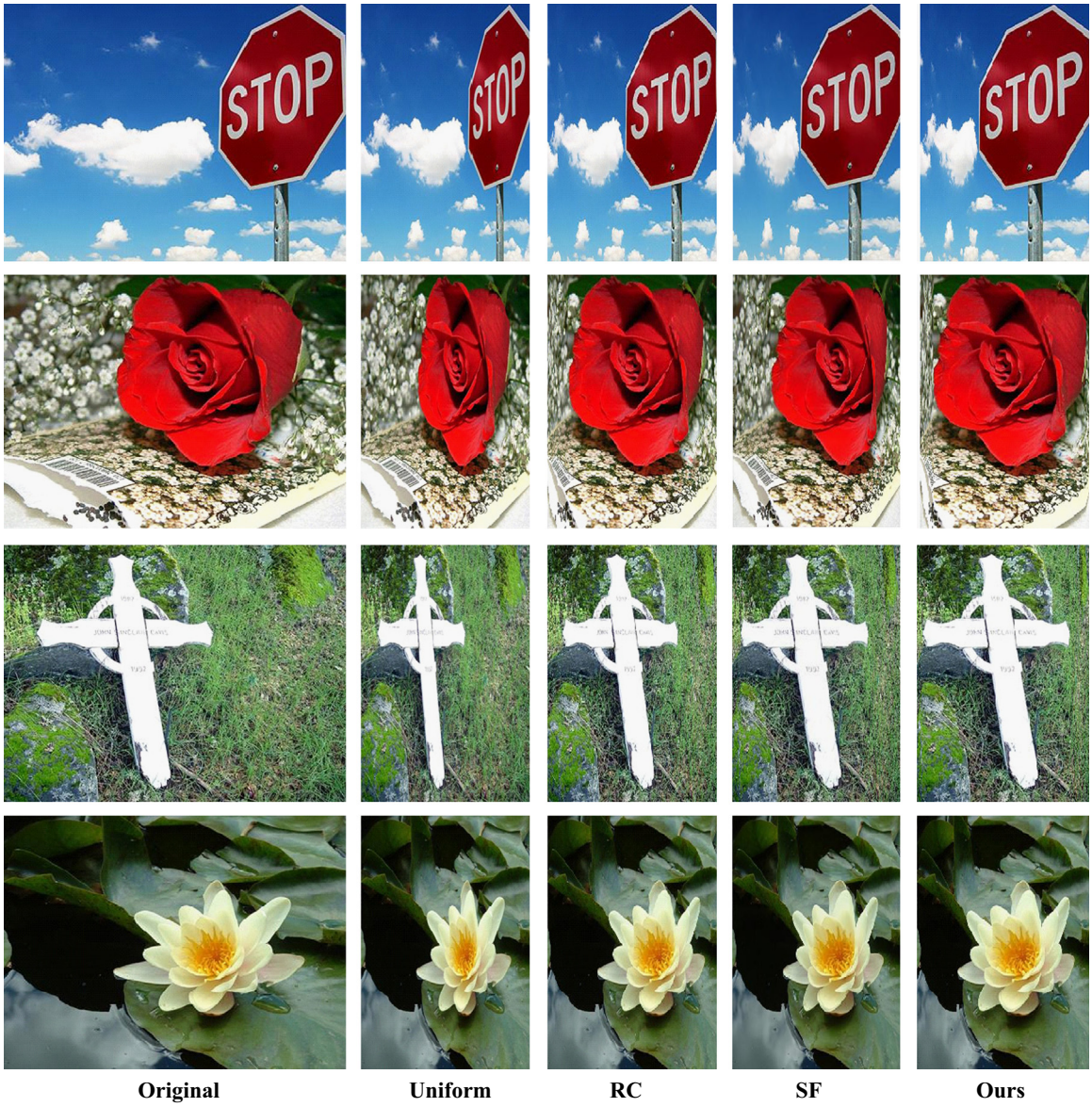

Fig. 17. Using saliency maps of RC, SF and ours on content-aware image resizing.

Fig. 18. Failure cases.

4.7. Computation cost

It takes about 2.7 s for our method to process a typical 400×300 color images. The most time consuming step is the pre-segmentation (SLIC) and refinement (mean-shift segmentation), which respective take about 1 s (37%) and 1.6 s (59.2%). The superpixel-based computation such as (4), (6) and (8) only takes 0.1 s (3.8%) in total. This is because (4), (6) and (8) all require the color distance between arbitrary superpixels. Actually the color distance between two superpixels only needs to be computed once when we first calculate (4) and then it could be reused for (6) and (8). The computation time reported above is acquired on our Dual Core 2.6 GHz laptop with 3 GB RAM using unoptimized Matlab code. Note we use Matlab wrapper of mex-file for SLIC and mean-shift segmentation.

4.8. Failure cases

As we use color information, we find our method fails in cases where object's colors and background colors are similar, as is shown in the first row of Fig. 18. Moreover, sometimes such low level feature may not really characterize the object that attracts human attention. In the second row of Fig. 18, the bag carried by the seated old man is rendered highest saliency. However, the real salient object should be the whole human body, not only the bag. This indicates that high-level features (top-down knowledge) are needed for a complete saliency detection system.

5. Conclusion

Our method effectively combines color contrast and distribution into a computational superpixel-based framework to meet the three disciplines for salient objects and renders high quality saliency maps. The exploited distribution prior and saliency smoothing procedure are both proved to contribute to the final results and achieve improvement in a large margin. Visual comparisons on the most popular dataset have shown the advantage of our method against other state-of-the-art approaches on popping out salient objects while suppressing the background. Evaluations under Precision–Recall and F-measure as well as application on content-aware image resizing have provided further support to the effectiveness of the proposed system. Since in this paper, only the color issues are considered, we may extent our future work towards multiple features and conduct more tests on other datasets.

Acknowledgment

This research is partly supported by National Science Foundation, China (no: 61273258, 61105001), Ph.D. Programs Foundation of Ministry of Education of China (no. 20120073110018).

References

[1] U. Rutishauser, D. Walther, C. Koch, P. Perona, Is bottom-up attention useful for object recognition? in: CVPR, 2004.
[2] J. Han, K. Ngan, M. Li, H. Zhang, Unsupervised extraction of visual attention objects in color images, IEEE Transactions on Circuits and Systems for Video Technology 16 (1) (2006) 141–145.
[3] T. Chen, M. Cheng, P. Tan, A. Shamir, S. Hu, Sketch2photo: internet image montage, ACM Transactions on Graphics 28 (5) (2009). 124:1–10.
[4] Y. Ding, X. Jing, J. Yu, Importance filtering for image retargeting, in: CVPR, 2011.
[5] Y. Pritch, E. Kav-Venaki, S. Peleg, Shift-map image editing, in: ICCV, 2009, pp. 151–158.
[6] M. Grundmann, V. Kwatra, M. Han, I. Essa, Discontinuous seam carving for video retargeting, in: CVPR, 2010.
[7] L. Wolf, M. Guttmann, D. Cohen-Or, Non-homogeneous content driven video-retargeting, in: ICCV, 2007.
[8] L. Itti, C. Koch, E. NieBur, A model of saliency-based visual attention for rapid scene analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (11) (1998) 1254–1259.
[9] D. Parkhurst, K. Law, E. Niebur, Modeling the role of salience in the allocation of overt visual attention, Vision Research 42 (1) (2002) 107–123.
[10] W. Einhauser, P. Konig, Does luminance-contrast contribute to a saliency map for overt visual attention? European Journal of Neuroscience 17 (5) (2003) 1089–1097.
[11] R. Fergus, P. Perona, A. Zisserman, Object class recognition by unsupervised scale-invariant learning, in: CVPR, 2003.
[12] J. Yang, M. Yang, Top-down visual saliency via joint CRF and dictionary learning, in: CVPR, 2012.
[13] T. Judd, K. Ehinger, F. Durand, A. Torralba, Learning to predict where humans look, in: ICCV, 2009.
[14] A. Borji, Boosting bottom-up and top-down visual features for saliency estimation, in: CVPR, 2012.
[15] A. Borji, D. Sihite, L. Itti, Probabilistic learning of task-specific visual attention, in: CVPR, 2012.
[16] M. Cheng, G. Zhang, N. Mitra, X. Huang, S. Hu, Global contrast based salient region detection, in: CVPR, 2011.
[17] F. Perazzi, P. Krahenbul, Y. Pritch, A. Hornung, Saliency filters: contrast based filtering for salient region detection, in: CVPR, 2012.
[18] X. Hou, L. Zhang, Saliency detection: a spectral residual approach, in: CVPR, 2007.
[19] Y. Zhai, M. Shah, Visual attention detection in video sequences using spatiotemporal cues, ACM in: Multimedia, 2006, pp. 815–824.

[20] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in: CVPR, 2009.
[21] S. Goferman, L. Zelnik-Manor, A. Tal, Context-aware saliency detection, in: CVPR, 2010.
[22] C. Rother, V. Kolmogorov, A. Blake, "Grabcut" – interactive foreground extraction using iterated graph cuts, ACM Transactions on Graphics 23 (3) (2004) 309–314.
[23] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, H. Shum, Learning to detect a salient object, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (2) (2011) 353–367.
[24] B. Alexe, T. Deselaers, V. Ferrari, What is an object?, in: CVPR, 2010, pp. 73–80.
[25] J. Feng, Y. Wei, L. Tao, C. Zhang, J. Sun, Salient object detection by composition, in: ICCV, 2011, pp. 1028–1035.
[26] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Ssstrunk, SLIC superpixels compared to state-of-the-art superpixel methods, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (11) (2012) 2274–2282.
[27] A. Borji, L. Itti, Exploiting local and global patch rarities for saliency detection, in: CVPR, 2012.
[28] D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (5) (2002) 603–619.
[29] C. Jung, C. Kim, A unified spectral-domain approach for saliency detection and its application to automatic object segmentation, IEEE Transactions on Image Processing 21 (3) (2012) 1272–1283.
[30] V. Gopalakrishnan, Y. Hu, D. Rajan, Salient region detection by modeling distributions of color and orientation, IEEE Transactions on Multimedia 11 (5) (2009) 892–905.
[31] Y. Fang, Z. Chen, W. Lin, C. Lin, Saliency detection in the compressed domain for adaptive image retargeting, IEEE Transactions on Image Processing 11 (5) (2012) 3888–3901.
[32] C. Lang, G. Liu, J. Yu, S. Yan, Saliency detection by multitask sparsity pursuit, IEEE Transactions on Image Processing 21 (3) (2012) 1327–1337.
[33] H. Li, K. Ngan, A co-saliency model of image pairs, IEEE Transactions on Image Processing 20 (12) (2011) 3365–3375.
[34] Z. Ren, Y. Hu, L. Chia, D. Rajan, Improved saliency detection based on superpixel clustering and saliency propagation, in: ACM International Conference on Multimedia, 2010.
[35] A. Borji, L. Itti, State-of-the-art in visual attention modeling, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (1) (2013) 185–207.
[36] A. Borji, D. Sihite, L. Itti, Salient object detection: a benchmark, in: ECCV, 2012.
[37] V. Movahedi, J. Elder, Design and perceptual validation of performance measures for salient object segmentation, in: IEEE Computer Society Workshop on Perceptual Organization in Computer Vision, 2010.
[38] L. Xu, H. Li, L. Zeng, King Ngan, Saliency detection using joint spatial-color constraint and multi-scale segmentation, Journal of Visual Communication and Image Representation 24 (4) (2013) 465–476.
[39] K. Fu, C. Gong, J. Yang, Y. Zhou, Salient object detection via color contrast and color distribution, in: ACCV, 2012.