Salient Object Detection via Saliency Spread

Dao Xiang, Zilei Wang

Department of Automation, University of Science and Technology of China,
Hefei, Anhui, 230027, China.

Abstract. Salient object detection aims to localize the most attractive objects within an image. For such a goal, accurately determining the saliency values of image regions and keeping the saliency consistency of interested objects are two key challenges. To tackle the issues, we first propose an adaptive combination method of incorporating texture with the dominant color, for enriching the informativeness and discrimination of features, and then propose saliency spread to encourage the image regions of the same object producing equal saliency values. In particular, saliency spread propagates the saliency values of the most salient regions to their similar regions, where the similarity serves for measuring the degree of belonging to the same object of different regions. Experimental results on the benchmark database MSRA-1000 show that our proposed method can produce more consistent saliency maps, which is beneficial to accurately segment salient objects, and is quite competitive compared with the advanced methods in previous literatures.

1 Introduction

Cognitive psychology research [1] indicates that given a visual scene, human vision is guided to particular parts by selective attention mechanism. These parts are called salient regions, and their saliency degree mainly depends on the state or quality of standing out from their neighbors. In computer vision, visual saliency simulates the functionality of selective attention, and concretely localizes the most salient and attention-grabbing regions or pixels in a digital image. Specifically, the saliency map represents the likelihood of each pixel belonging to salient regions with different values. Visual saliency estimation is much helpful to various vision tasks, such as object detection and recognition [2, 3], adaptive image display [4], content-aware image editing [5], and image segmentation [6–8]. Recently, besides the eye-fixation prediction, visual saliency begins to serve object detection with the aim of segmenting salient objects from images. Particularly, this work focuses on such a detection goal.

Inspired by the pioneering work in [9], different saliency models for detecting salient objects were proposed. Most of them [10, 11, 8, 7] use the superpixel-level color contrast to compute saliency map, due to the special attention of human vision to color and the robustness of superpixels compared with raw pixels [12]. However, these methods unavoidably suffer from unsatisfied segmentation results, *i.e.*, either producing incomplete objects or being contaminated by background. In our opinion, the reasons of leading to such unexpected results are two

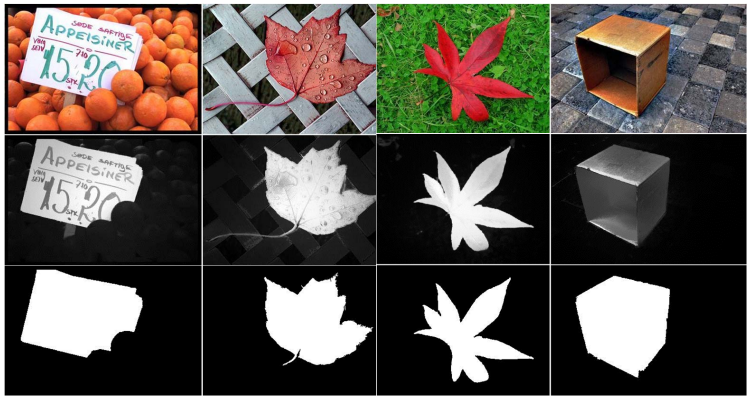

Fig. 1. Illustration of the effectiveness of our method in detecting salient objects. From top to bottom: input images, saliency maps obtained by our method, and our segmentation results.

fold. The first is the insufficiency of color feature. Only adopting color feature works well for most natural images with considerable color variance between foreground and background, but not for the images without dominant color yet (*e.g.*, artificial images or gray-scale images). Consequently, the poor segmentations are produced (see Fig. 2) as little information is provided. Thus more visual cues need to be incorporated. Along this routine, some improved methods [13–15] have been proposed with different combination means of multiple features. The second is the inconsistency of the saliency of object regions. Under a certain saliency model, different parts of the ground truth salient object are likely not to produce uniform saliency, due to the object internal incoherence and the model sensitivity [16]. So different pixels or superpixels of the same object would have inconsistent saliency values. And such saliency map would result in the failure of exactly keeping the completeness of the segmented objects without absence of object parts or contamination of background (see Fig. 3). This fact is actually an important challenge of detecting salient objects.

To alleviate the aforementioned issues, we propose two concrete approaches in this paper to improve the performance of saliency detection. Firstly, we propose an adaptive feature fusion strategy for incorporating the texture with the main color feature.

Secondly, we propose a **saliency spread** mechanism to tackle the saliency inconsistency of object regions. The main idea of saliency spread is to spread saliency values of the most salient regions with high confidence to the similar regions by exploring the feature correlation of regions (probably belonging to the same object).

Figure 1 gives some examples of our proposed method to segment objects. Before elaborating on the details of our method, we review the related works

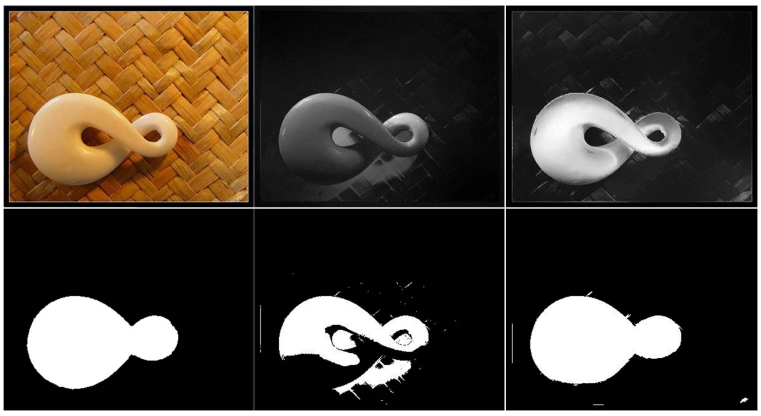

Fig. 2. Exemplar of the insufficiency of color features. For images whose foreground and background have similar color distributions, the method (SF [7] here) only using color leads to pool performance (middle column), while our method can achieve much better segmentation result due to incorporating texture (right column).

on detecting the salient objects. More detailed investigation and comparing can also be found in [16].

The rest of this paper is organized as follows. We first review the related works of saliency object detection in Section 2. Then we give an overview of our method in Section 3, and the detailed description of key models in Section 4. Finally, in Section 5, we report experimental results of the proposed method on public benchmark. The conclusions are provided in Section 6.

2 Related Works

In this paper, we focus on the data-driven bottom-up saliency detection. This kind of saliency is usually derived by primitive image features, such as color, texture, and edges. Based on the design ideology, the bottom-up saliency detection methods can roughly be classified into three categories: (1) frequency domain analysis based methods: the saliency is determined by the amplitude or phase spectrum [17, 18]; (2) information theory based methods: Shannon's self-information [19] or the entropy of the sampled visual features [20] is maximized for achieving attention selectivity; (3) contrast based methods: the saliency map is computed by exploring the contrast of image pixels or regions. Now we briefly review the contrast based methods since this work falls into this one.

Actually, the contrast based methods have been proved to achieve the state-of-the-art performance [8, 7, 10, 11, 21–23]. Perceptual research results [24, 25] indicate that contrast is the most influential factor in low-level stimuli-driven

Fig. 3. Exemplar of the saliency inconsistency of object regions. For a salient object without uniform color distributions, the traditional methods (SF [7] here) fail to exactly segment the complete object (middle column), while our method with saliency spread can significantly improve the quality of segmentation (right column).

attention. Itti *et al.* proposed the fundamental framework of the contrast model [9], which particularly uses center-surrounded differences across multi-scale low-level features to detect saliency. A typical workflow of such methods includes extracting multiple low-level features (color, intensity, orientation, etc) to construct prominent maps by determining the contrast of image regions to their surroundings, and combining these maps to form a final saliency map via a predefined fusion strategy.

The contrast based methods can use local or global information. The local contrast based methods utilize the neighborhoods to estimate the saliency of a certain image region. For example, Liu *et al.* [10] defines multi-scale contrast as a linear combination of contrasts in a Gaussian image pyramid. Ma *et al.* [23] generates a saliency map based on dissimilarities at the pixel-level, and extracts attended areas or objects using a fuzzy growing method. These local contrast based methods tend to highlight the object boundaries rather than the entire area, which limits the segmentation-like applications. In contrary, the global contrast based methods consider the contrast relations within the whole image to evaluate saliency of an image region. Zhai *et al.* [26] defines pixel-level saliency based on a pixel's contrast to all other pixels. Chen *et al.* [8] simultaneously evaluates global contrast differences and spatial coherence. Perazzi *et al.* [7] computes two kinds of contrasts (*i.e.*, uniqueness and the spatial distribution) of perceptually homogeneous regions with weighting parameters to compromise local and

global contrast. Though these global models achieve more consistent results, they may fail to highlight the entire target objects, or get rid of background. In this work, we impute these inferiors to the insufficiency of color feature and the inconsistency of the saliency of object regions. Specifically, we propose two strategies to improve the saliency detection performance from the feature fusion and the saliency consistency.

For enriching the informativeness of features, we consider the texture to serve as a supplementation of color feature. In the previous literatures, the texture has actually been used for providing the information of spatial arrangement of color or intensities. Tang *et al.* [13] incorporates the LBP texture into color for providing diverse information, and the combined features can achieve a better saliency detection performance. Gopalakrishnan *et al.* [14] simultaneously computes the color saliency map and the orientation saliency map, then chooses the one of higher connectivity and less spatial variance as the final saliency map. However, these methods suffer from either model complexity or failing to find the accurate object boundary. In this work, we specifically use the LM filter bank [27] to produce the texture feature, and combine it with the color in an adaptive manner, which depends on the image content.

As for the inconsistency of salient object parts incurred by the model sensitivity [16], we propose saliency spread to alleviate it. Here we assume that different regions of the same object have similar color or texture distribution. So we can utilize the correlation of object parts to encourage the similar parts (likely belonging to the same object) producing equal saliency value. Specifically, we first pick out the most salient regions, and then use the relationship with these regions to enhance the saliency of similar regions, where the similarity of regions is determined by their color, texture, and position in practice. To the best of our knowledge, no similar works have been proposed yet.

3 Overview

In this section, we briefly introduce the framework of our proposed method. We follow the classical pipeline of the contrast based methods except that the proposed saliency spread is embedded. Therefore, as shown in Figure 4, our method is composited of four key stages: (1) generating the superpixels of images as homogeneous regions, (2) computing the saliency values of image regions, (3) conducting saliency spread to highlight the saliency object, and (4) assigning each pixel a saliency value to produce the final saliency map.

3.1 SuperPixel Generation

This step is used to decompose an image into superpixels [28], which are small regions grouped by homogenous neighboring pixels with similar properties (color, brightness, texture, etc.). The superpixel-level saliency estimation is more robust and efficient than the pixel-level one in practice [8, 7]. In fact, superpixels could capture image redundancy and abstract unnecessary details, which conforms

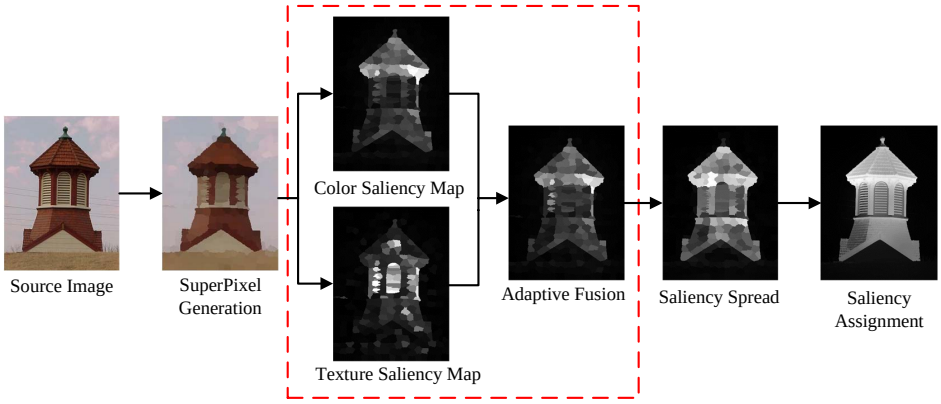

Fig. 4. The framework of our proposed saliency spread method, which includes superpixel generation, regional saliency computation, saliency spread, and pixel-level saliency assignment. Particularly, saliency spread encourages the saliency values of regions belonging to the same objects to be consistent.

with the regional perception mechanism of human vision. Moreover, superpixels could significantly decrease the number of involved elements, which will reduce the computational complexity.

In this work, we adopt the SLIC method [29] to decompose an image into superpixels, which are denoted by $R = R_1, R_2, ..., R_M$. Specifically, SLIC employs K-means clustering to segment images in the CIELab color space, and consequently the compact, memory efficient and edge-preserving superpixels can be yielded.

3.2 Regional Saliency Computation

This stage is used for computing the saliency value of each region produced in the first step. Generally, the saliency of one region is determined by the properties of itself and the contrast relationship with its neighbors. From such considerations, we use two kinds of features (color and texture) for enriching the description, and define two contrast metrics, *i.e.*, uniqueness and distribution, to measure the saliency.

Here *uniqueness* represents the rarity or surprise of a region, which has actually been used for saliency detection in the previous works [8, 7]. Such definition is natural to saliency computation since the regions with unusual surroundings are more attractive for human vision. In this work, we propose a revised version of such uniqueness by considering more information. *Distribution* denotes the spatial variances of features within a certain region. Roughly speaking, the distribution of features belonging to the foreground is probably more centralizing, while for the background it may exhibit more diverse with high spatial variance [30, 10, 7].

3.3 Saliency Spread

Most existing saliency estimation models directly obtain the final saliency map from the fused saliencies of multiple features computed in the previous stage. Different from them, we propose *saliency spread* to enhance the consistency of different regions of salient objects. In practice, it is observed that the saliency values of object regions can vary seriously due to the model sensitivity [16], even they have similar color and texture (see Figure 4), or the feature variation of object parts. Saliency spread tried to tackle the issue by utilizing the similarity of image regions (probably belonging to the same object), and can be regarded as a special smoothing technique on regional saliency. Specifically, we first pick out the n most salient regions as a pseudo-object, and then enhance similar regions via propagating the saliency of the selected pseudo-object, where the color, texture, spatial position are comprehensively leveraged.

3.4 Saliency Assignment

The role of this step is to assign each pixel a saliency value using regional saliencies. Directly assigning each pixel the same value as the belonged region would lose detailed information within superpixels(*e.g.*, strong edges or small feature variations), and thus much error is caused. So we adopt the upsampling method used in [7], which works well due to the ability of capturing details and preserving edges.

4 Algorithm

In the following section we will give a detailed description of regional saliency computation and saliency spread, which form the main parts of our method.

4.1 Regional Saliency Computation

In this section, we show in detail how to measure the two kinds of contrast, *i.e.*, uniqueness and distribution, for color and texture respectively, and combine the power of them to generate the final regional saliency map.

Uniqueness As mentioned before, uniqueness generally stands for the rarity of a region with its surroundings. Hence the key issues in uniqueness are to determine *surroundings* and characterize *rarity*. *Surroundings* represents the regions involved in computing *rarity*, which should have nonuniform significance due to their spatial positions. And *rarity* denotes the regional feature difference. Intuitively, distance is a proper choice to measure both of them. So we naturally give

the definition of *color uniqueness* for R_i:

$$\begin{aligned} U_i^c &= \sum_{j=1}^{M} r_j \cdot d_{i,j}^c \cdot d_{i,j}^p \\ d_{i,j}^c &= \chi^2(c_i, c_j) = \sum_{k=1}^{t} \frac{(h_{1k} - h_{2k})^2}{h_{1k} + h_{2k}} \\ d_{i,j}^p &= \exp(-\frac{1}{2\sigma_u^2} \cdot \|p_i - p_j\|_2^2) \end{aligned} \tag{1}$$

where r_j is the number of pixels in R_j, and emphasizes the contrast to bigger regions. $d_{i,j}^c$ is the chi-square distance between color histograms of R_i and R_j. c_i is the color histogram of R_i in Lab colorspace with $t = 60$ bins. A small variation of a or b channel could cause a remarkable change of color perception when they are close to 0, so we non-uniformly quantize a and b to 22 bins respectively with well chosen quantization near 0. To be more specific, the quantization intervals of a and b below 0 are set as follows: $[-127, -70]$, $(-70, -60]$, $(-60, -50]$, $(-50, -40]$, $(-40, -30]$, $(-30, -25]$, $(-25, -20]$, $(-20, -15]$, $(-15, -10]$, $(-10, -5]$, $(-5, 0]$. Symmetrically, the quantization density of a and b above 0 stays the same with the one below 0. We choose color histogram to alleviate the information loss of using mean color [7] or algorithm complexity caused by exhaustively computing distances among all the colors in R_i and R_j [8].

$d_{i,j}^p$ represents the spatial relationship between R_i and R_j, and renders R_j as more important when they're close. p_i is the mean position of R_i. The introduction of $d_{i,j}^p$ can effectively compromise the global and local contrast, allowing for a sensitivity to local color variation and meanwhile avoiding overemphasizing object edges. Actually in extreme cases, where $d_{i,j}^p = 1$, (1) is equivalent to a completely global uniqueness estimation [8], whereas $d_{i,j}^p \approx 0$ if R_i and R_j are not direct neighbors will yield a local contrast estimation [10]. Parameter σ_u tunes the range of the uniqueness operator. In practice, we find that $\sigma_u = 0.15$ is a well tuned value.

Similar to (1), we define *texture uniqueness* as:

$$\begin{aligned} U_i^t &= \sum_{j=1}^{M} r_j \cdot d_{i,j}^t \cdot d_{i,j}^p \\ d_{i,j}^t &= \|t_i - t_j\|_2^2 \end{aligned} \tag{2}$$

where t_i is the texture feature of R_i. Here we use the max response among the LM filter bank [27] to represent t_i. The LM set is a multi-scale, multi-orientation filter bank with 48 filters, which consists of first and second derivatives of Gaussians at 6 orientations and 3 scales making a total of 36, 8 Laplacian of Gaussian (LOG) filters, and 4 Gaussians.

With the above definitions, we combine the power of color and texture to get a enhanced *uniqueness* of R_i:

$$U_i = w \cdot U_i^c + (1 - w) \cdot U_i^t \tag{3}$$

where w depends on the image. The contribution of color and texture differs across images. Hence it's not suitable to use a fixed value as the weight. Noticing that the more information the color or texture provides, the greater is its uniqueness variation, we use uniqueness variation to represent the contribution of color and texture. To be more specific, we set $w = \xi \cdot \mathrm{var}(U_c)/(\xi \cdot var(U_c) + \mathrm{var}(U_t))$, where ξ is a tuning parameter to highlight the importance of color, and $var(*)$ represents the variation. A similar idea can be found in [31], where the weights of color and texture are determined by computing the overlapping degree of their distributions given the foreground and background sample. In all our experiments, we set $\xi = 5$.

Distribution Features belonging to the foreground are generally compact and exhibit low spatial variances. So we define regional distribution using the spatial variances of its features. The spatial variance of a feature corresponds to its occurrence elsewhere in the image, which can be measured by its spatial distance to the mean position. Thus we define *color distribution* for R_i as:

$$\begin{aligned} D_i^c &= \sum_{j=1}^{M} r_j \cdot \|p_j - p_i^c\|_2^2 \cdot \tilde{d}_{i,j}^c \\ p_i^c &= \sum_{j=1}^{M} r_j \cdot \tilde{d}_{i,j}^c \cdot p_j \\ \tilde{d}_{i,j}^c &= \exp(-\frac{1}{2\sigma_d^2}\chi^2(c_i, c_j)) \end{aligned} \tag{4}$$

where p_i^c is the weighted mean position of R_i in terms of color. $\tilde{d}_{i,j}^c$ denotes color similarity between R_i and R_j, which is defined with color distance. The parameter σ_d controls the role that color similarity plays, since a big σ_d tends to decrease the significance of regions with similar color, while a small one yields more sensitivity to color variation. In our experiments, we set $\sigma_d = 10$.

The *texture distribution* is defined in a similar way to (4):

$$\begin{aligned} D_i^t &= \sum_{j=1}^{M} r_j \cdot \left\|p_j - p_i^t\right\|_2^2 \cdot \tilde{d}_{i,j}^t \\ p_i^t &= \sum_{j=1}^{M} r_j \cdot \tilde{d}_{i,j}^t \cdot p_j \\ \tilde{d}_{i,j}^t &= \exp(-\frac{1}{2\sigma_d^2}\left\|t_i - t_j\right\|_2^2) \end{aligned} \tag{5}$$

where t_i is again the texture feature. We combine color and texture distribution with adaptive weighting to obtain the *distribution* of region R_i:

$$D_i = w \cdot D_i^c + (1 - w) \cdot D_i^t \tag{6}$$

Saliency Fusion After obtaining uniqueness U_i and distribution D_i for region R_i, we now combine them to obtain a regional saliency map. Assuming that U_i and D_i are independent, we define the saliency value S_i^f of region R_i similar to [7]:

$$S_i^f = U_i \cdot \exp(-\lambda \cdot D_i) \tag{7}$$

The form of exponential function is chosen to emphasize D_i, which is more powerful to highlight salient regions. The scaling factor λ is empirically set to 3 in our experiments.

4.2 Saliency Spread

This step is to deal with the inconsistent saliencies of object parts.We assume that regions belonging to the same objects have similar properties, and choose n most salient regions as pseudo-objects, then spread saliency to the regions that are likely to belong to the selected objects. This can be formulated as:

$$S_i = S_i^f + \sum_{j=1}^{n} r_j \cdot S_j^f \cdot \exp(-\frac{\chi^2(c_i, c_j)}{2\alpha^2} - \frac{\|t_i - t_j\|_2^2}{2\beta^2} - \frac{\|p_i - p_j\|_2^2}{2\delta^2}) \tag{8}$$

where S_i^f is the saliency value of region S_i obtained from (7). α, β, δ are tuning parameters that adjust the significance of color, texture and spatial relations with selected pseudo-object regions, respectively. In our experiments, we set $\alpha = \beta = \delta = 2$, which is strong enough to guarantee that only the nearby regions with similar color and texture are enhanced. From figure 4 we can see that saliency spread could significantly increase the saliency value of the object regions and highlight the object as a whole. In our experiment, we empirically set $n = 30$.

Saliency spread can bring another benefit. Many methods are based on the assumption that the containing objects in an image are in a position near the center of the image. [10] use the distance from pixel x to the image center as a weight to assign less importance to colors nearby image boundaries, [11] treat the 15-pixel wide narrow border region of the image as pseudo-background region and extract the backgroundness descriptor. But such assumption is not always true, and the methods will have a poor performance on images in which objects reside near the boundaries. Our saliency spread could roughly determine the location of objects without the assumption, and this will help to relatively decrease the backgroundness saliency via (8).

The last step is a per-pixel saliency assignment. For pixel i:

$$Sal_i = \sum_{j=1}^{M} r_j \cdot S_j \cdot \exp(-\frac{1}{2\sigma_c^2} \cdot \chi^2(c_i, c_j) - \frac{1}{2\sigma_p^2} \cdot \|p_i - p_j\|_2^2) \tag{9}$$

where S_j are regional saliency surrounding pixel i. σ_c and σ_p are parameters controlling the sensitivity to color and position respectively, we set $\sigma_c = \sigma_p = \frac{1}{30}$ in the experiments. Finally, the resulted pixel-level saliency map is rescaled to the range $[0-255]$ for the purpose of exhibiting and comparing with the groundtruth.

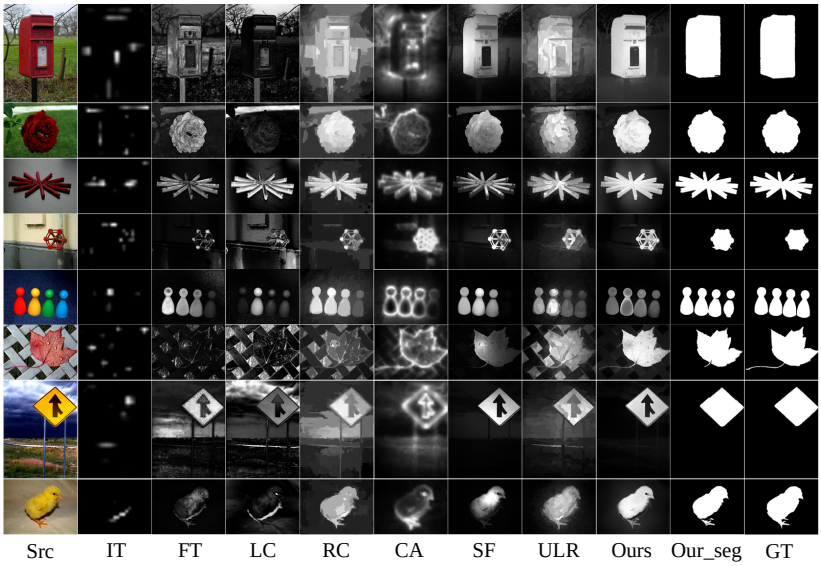

Fig. 5. Visual comparison of previous approaches to our method. Due to space limitation, only a part of the results are exhibited. Our method generates consistent and uniform salient regions. The segment results (Ours-Seg), which are obtained using adaptive threshold (Eq.10), are also close to ground truth (GT).

5 Experiments

We evaluate the results of our approach on the commonly used MSRA-1000 databset provided by [32], which is a subset of MSRA [10]. MSRA-1000 is the largest of its kind [8] for saliency detection with accurate human-marked labels as binary ground truth rather than rectangle bounding boxes used in MSRA. We provide a comprehensive comparison of our method to 13 state-of-the-art saliency detection methods, including biologically-motivated saliency (IT [9]), purely computational fuzzy growing (MZ [23]), frequency domain based saliency (FT [32], SR [18]), spatiotemporal cues (LC [26]), graphed-based saliency (GB [33]), context-aware saliency (CA [30]), salient region detection (AC [21]), low-rank matrix recovery theory inspired saliency (LSMD [34], ULR [35]), and works related to our method (SF [7], HC [8], RC [8]). To evaluate these methods, we use author's implementation (when available) or the resulting saliency maps provided in [8]. A visual comparison of saliency maps obtained by these methods can be seen in Figure 5.

In order to comprehensively evaluate the performance of our method, we conduct two experiments following the standard evaluation measures in [7, 34, 8]. In the first experiment, we segment saliency maps using fixed or adaptive threshold, and calculate precision and recall curves. In the second experiment,

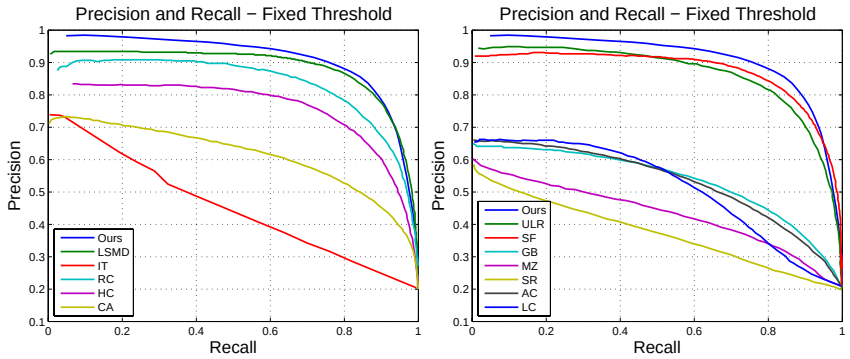

Fig. 6. Precision-Recall curves for fixed threshold of saliency maps. Compared with various methods, our approach achieves the best performance.

we use mean absolute error to evaluate how well the continuous saliency map match the binary ground truth.

5.1 Segmentation with thresholding

A common way for assessing the accuracy of saliency detection methods is to binarize each saliency map with fixed threshold or adaptive threshold, and compute its precision and recall rate. Precision (also called positive predictive value) represents the fraction of retrieved pixels that are relevant, while recall (also known as sensitivity) corresponds to the percentage of relevant pixels that are retrieved. They are often evaluated simultaneously, since a high precision can be obtained at the cost of a low recall and vice-versa.

Fixed Threshold We first segment a saliency map with a fixed threshold $t \in [0, 255]$. After the segmentation, we compare the binarized image with ground truth to obtain its precision and recall. To reliably measure the capability of various methods highlighting salient regions in images, we vary the threshold t from 0 to 255 to generate a sequence of precision-recall pairs. After averaging over all the results of images in the dataset, we obtain the precision-recall curves, as Fig 6 shows. As we can see, compared to other approaches, the saliency maps generated by our method with fixed threshold are more accurate, and closer to the ground truth on the whole.

Adaptive Threshold Similar to [7, 35], we adopt the image dependent adaptive threshold, which is defined as twice the mean saliency value of the entire image [32]:

$$T_a = \frac{2}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} S(x, y) \tag{10}$$

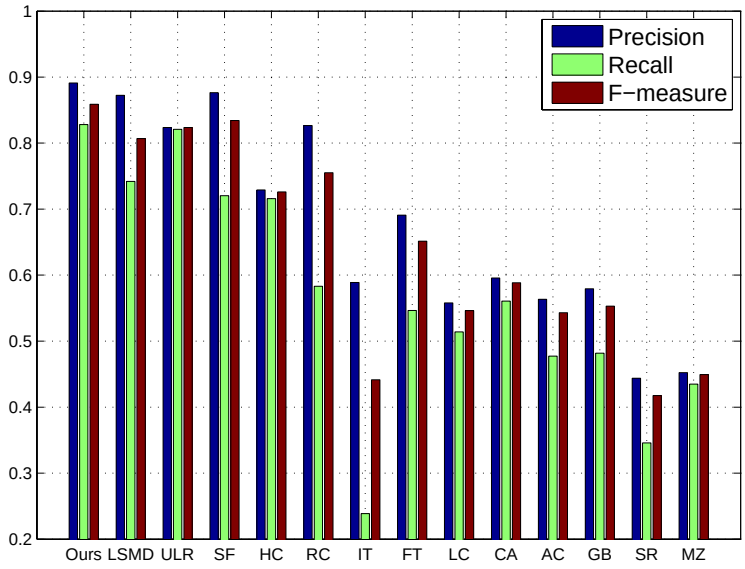

Fig. 7. Precision, recall, and F-measure for adaptive thresholds.

where W and H are the width and height of the image, respectively. S is the obtained saliency map. Adaptive threshold is a simple but practical indicator for comparing quality among approaches, as the resulting segmentation could be directly utilized in other literatures. In addition to precision and recall, we also compute their weighted harmonic mean measure (or F-measure), which is defined as:

$$F_\beta = \frac{(1+\beta^2)\cdot \text{Precision}\cdot \text{Recall}}{\beta^2 \cdot \text{Precision} + \text{Recall}} \tag{11}$$

Similar to previous works [32, 7], we also set $\beta^2 = 0.3$. The result is given in Figure 7. Our method achieves the best precision, recall and F-measure among all the approaches. Compared to SF, which is closest to our method, we have a significant improvement of recall (9%), which means our method are likely to detect more salient regions, while keeping a high accuracy.

5.2 Mean Absolute Error

Ideally a saliency map should be equal to the ground truth, and each thresholding in $(0, 255)$ results in the same segmentation, *i.e.* the true object. Hence the more similar with the ground truth, the better is the saliency map and the algorithm generating it. Yet neither the precision nor recall measure consider such performance indicator. We adopt MAE (Mean Absolute Error) to measure the similarity between the continuous saliency map S and the binary ground

truth GT, which is defined in [7]:

$$\mathrm{MAE} = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} |S(x,y) - GT(x,y)| \tag{12}$$

where W and H are again the width and the height of the respective saliency map and ground truth image. We compute MAE by averaging over all images with the same parameter settings. Figure 8 shows that our method generates the lowest MAE measure, which means that our saliency maps are more consistent with the ground truth.

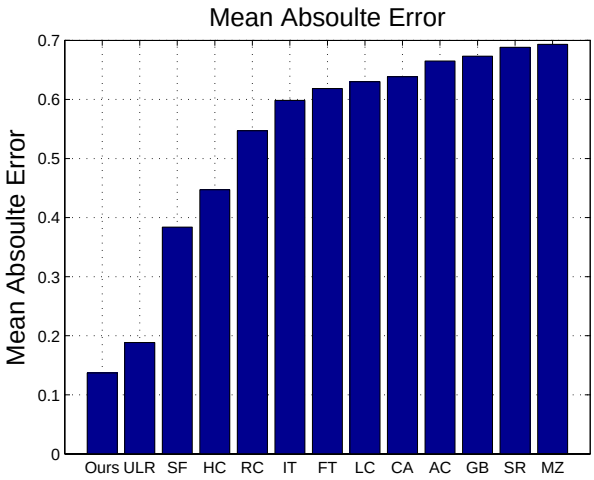

Fig. 8. Mean absolute error of the different saliency methods to ground truth.

6 Conclusions

In this work, we present a contrast based method for salient object detection, which follows the typical pipeline of contrast measures estimating and fusing. On this basis, we analysis the weakness of existing models, and attribute it to the insufficiency of color feature and the inconsistency of the saliency of object regions. Contrapose these deficiencies we present two improvements. Firstly, we incorporate texture as a complementary feature of color, to deal with images without dominant color. Secondly, we propose saliency spread, which propagates saliencies to regions that are likely to belong to the same objects and achieves more consistent saliency maps. Experiments show the superiority of our proposed schemes in terms of serval widely accepted indicators.

References

1. Mangun, G.R.: Neural mechanisms of visual selective attention. Psychophysiology **32** (1995) 4–18
2. Kanan, C., Cottrell, G.: Robust classification of objects, faces, and flowers using natural image statistics. In: CVPR, IEEE (2010) 2472–2479
3. Rutishauser, U., Walther, D., Koch, C., Perona, P.: Is bottom-up attention useful for object recognition? In: CVPR. Volume 2. (2004) II–37
4. Chen, L.Q., Xie, X., Fan, X., Ma, W.Y., Zhang, H.J., Zhou, H.Q.: A visual attention model for adapting images on small displays. Multimedia systems **9** (2003) 353–364
5. Ding, M., Tong, R.F.: Content-aware copying and pasting in images. The Visual Computer **26** (2010) 721–729
6. Ko, B.C., Nam, J.Y.: Object-of-interest image segmentation based on human attention and semantic region clustering. JOSA A **23** (2006) 2462–2470
7. Perazzi, F., Krahenbuhl, P., Pritch, Y., Hornung, A.: Saliency filters: Contrast based filtering for salient region detection. In: CVPR. (2012) 733–740
8. Cheng, M.M., Zhang, G.X., Mitra, N.J., Huang, X., Hu, S.M.: Global contrast based salient region detection. In: CVPR. (2011) 409–416
9. Itti, L., Koch, C., Niebur, E., et al.: A model of saliency-based visual attention for rapid scene analysis. PAMI **20** (1998) 1254–1259
10. Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., Shum, H.Y.: Learning to detect a salient object. PAMI **33** (2011) 353–367
11. Jiang, H., Wang, J., Yuan, Z., Wu, Y., Zheng, N., Li, S.: Salient object detection: A discriminative regional feature integration approach. In: CVPR, IEEE (2013) 2083–2090
12. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. IJCV **59** (2004) 167–181
13. Tang, K., Au, O.C., Fang, L., Yu, Z., Guo, Y.: Multi-scale analysis of color and texture for salient object detection. In: ICIP. (2011) 2401–2404
14. Gopalakrishnan, V., Hu, Y., Rajan, D.: Salient region detection by modeling distributions of color and orientation. Multimedia **11** (2009) 892–905
15. Hu, Y., Xie, X., Ma, W.Y., Chia, L.T., Rajan, D.: Salient region detection using weighted feature maps based on the human visual attention model. In: Advances in Multimedia Information Processing-PCM. Springer (2005) 993–1000
16. Borji, A., Sihite, D.N., Itti, L.: Salient object detection: A benchmark. In: ECCV. Springer (2012) 414–429
17. Guo, C., Ma, Q., Zhang, L.: Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In: CVPR. (2008) 1–8
18. Hou, X., Zhang, L.: Saliency detection: A spectral residual approach. In: CVPR. (2007) 1–8
19. Bruce, N., Tsotsos, J.: Saliency based on information maximization. Advances in neural information processing systems **18** (2006) 155
20. Hou, X., Zhang, L.: Dynamic visual attention: searching for coding length increments. In: NIPS. Volume 5. (2008) 7
21. Achanta, R., Estrada, F., Wils, P., Süsstrunk, S.: Salient region detection and segmentation. In: Computer Vision Systems. Springer (2008) 66–75
22. Duan, L., Wu, C., Miao, J., Qing, L., Fu, Y.: Visual saliency detection by spatially weighted dissimilarity. In: CVPR, IEEE (2011) 473–480

23. Ma, Y.F., Zhang, H.J.: Contrast-based image attention analysis by using fuzzy growing. In: Proceedings of the eleventh ACM international conference on Multimedia, ACM (2003) 374–381
24. Einhäuser, W., König, P.: Does luminance-contrast contribute to a saliency map for overt visual attention? European Journal of Neuroscience **17** (2003) 1089–1097
25. Parkhurst, D., Law, K., Niebur, E.: Modeling the role of salience in the allocation of overt visual attention. Vision research **42** (2002) 107–123
26. Zhai, Y., Shah, M.: Visual attention detection in video sequences using spatiotemporal cues. In: ACM Multimedia, ACM (2006) 815–824
27. Leung, T., Malik, J.: Representing and recognizing the visual appearance of materials using three-dimensional textons. IJCV **43** (2001) 29–44
28. Ren, X., Malik, J.: Learning a classification model for segmentation. In: Computer Vision, IEEE (2003) 10–17
29. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels. EPFL, Tech. Rep **2** (2010) 3
30. Goferman, S., Zelnik-Manor, L., Tal, A.: Context-aware saliency detection. PAMI **34** (2012) 1915–1926
31. Shahrian, E., Rajan, D.: Weighted color and texture sample selection for image matting. In: CVPR, IEEE (2012) 718–725
32. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: CVPR. (2009) 1597–1604
33. Harel, J., Koch, C., Perona, P., et al.: Graph-based visual saliency. Advances in neural information processing systems **19** (2007) 545
34. Peng, H., Li, B., Ji, R., Hu, W., Xiong, W., Lang, C.: Salient object detection via low-rank and structured sparse matrix decomposition. In: AAAI. (2013)
35. Shen, X., Wu, Y.: A unified approach to salient object detection via low rank matrix recovery. In: CVPR, IEEE (2012) 853–860