SALIENT REGION DETECTION VIA TEXTURE-SUPPRESSED BACKGROUND CONTRAST

Jiamei Shuai[1,2] *Laiyun Qing*[1,2] *Jun Miao*[2] *Zhiguo Ma*[2] *Xilin Chen*[2]

[1] University of Chinese Academy of Sciences, Beijing, 100049, China
[2] Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, China

ABSTRACT

We propose a novel salient region detection algorithm by texture-suppressed background contrast. We employ a structure extraction algorithm to suppress the small scale textures which are supposed to be not sensitive for human vision system. Then the texture-suppressed image is segmented into homogeneous superpixels. Motivated by the observation that the spatial distribution of the background has a high probability on the boundaries of images, we estimate the background as superpixels near the image boundaries. The saliency of each superpixel is then defined as the summation of its k minimum color distances to the estimated background superpixels. Finally a post-processing process involving spatial and color adjacency is employed to generate a per-pixel saliency map. Experimental results demonstrate that the proposed method outperforms the state-of-the-art approaches.

Index Terms— Salient region detection, Background contrast, Texture suppression, Superpixels

1. INTRODUCTION

Human vision system (HVS) can process massive visual information at a glance and fixate at salient objects in a scene. Extensive efforts have been devoted to the research of bottom-up saliency models to achieve equivalent functionality. Saliency estimation can be applied to many computer vision tasks, such as object detection [1], image segmentation [2], object tracking [3], etc.

One of the early works on saliency estimation is the bottom-up method proposed by Itti *et al.* [4], which determined visual saliency as center-surround contrast using a difference of Gaussians (DoG) approach. Gao *et al.* [5] utilized the mutual information between the feature distribution of center and surround regions to estimate saliency. These approaches [4, 5, 6] focus on predicting human fixations in natural images, rather than locating salient objects.

Recent works on saliency detection have paid more attention to salient object detection [7, 8, 9, 10, 11, 12, 13, 14]. Various criterions on measuring contrast and global rarity have been explored. Zhai *et al.* [13] defined pixel-level saliency based on the pixel's contrast to all other pixels of the image. Feng *et al.* [12] measured the saliency of an image window as the cost of composing the window using the remaining part of the image. Cheng *et al.* [9] produced region-based saliency maps using the region's contrast and spatial distances to other regions in the image. Perazzi *et al.* [10] proposed a contrast based filtering algorithm and defined superpixel-level saliency by combining two contrast measures namely element uniqueness and distribution.

Motivated by the observation that the image boundaries are mostly background, we estimate the background as the regions near image boundaries and define saliency as the contrast with the background, rather than globally with all the other regions. Similar idea has been utilized in Wei *et al.* [11], which defined the saliency of an image patch as its geodesic distance to the background. Different from [11], which relied on the connectivity prior of background, we explore the homogeneous nature of background and define the saliency of an image region as the summation of its k minimum color distances to the background regions. In the proposed algorithm, we adopt superpixel segmentation to generate homogeneous superpixels. Moreover, a pre-processing step is applied to suppress the small scale textures in images, since human vision system is not sensitive to such small variances when detecting salient locations.

The remainder of this paper is organized as follows: Section 2 introduces details of the proposed approach. Section 3 presents experiments and the conclusion follows in Section 4.

2. PROPOSED APPROACH

We propose a novel approach to detect salient regions in natural images through *Texture-suppressed Background contrast* (*TB* for short). The framework of the proposed approach is presented in Fig. 1. A structure extraction algorithm is applied to suppress the small variances in textures. Then the texture-suppressed image is segmented into superpixels. The background are estimated as the superpixels near image boundaries, according to the fact that image boundaries are mostly background. We define our per-superpixel saliency as its con-

This research is partially sponsored by National Basic Research Program of China (No.2009CB320900), and Natural Science Foundation of China (Nos.61070116, 61070149, 61001108, 61175115, and 61272320).

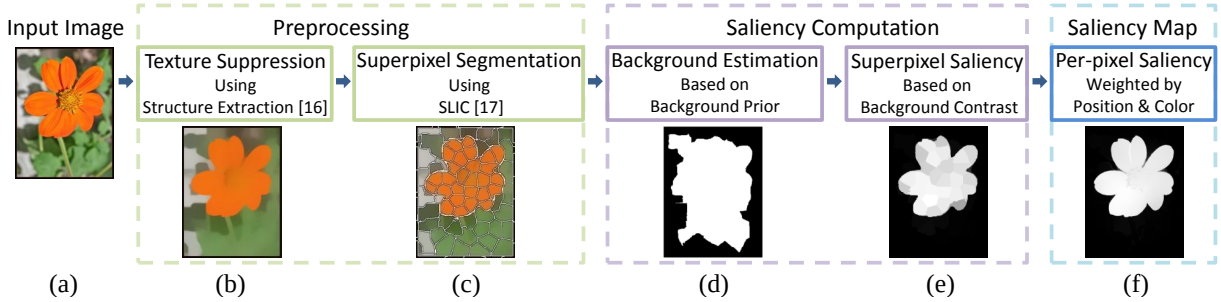

Fig. 1. Framework of our approach. (a) Input image (b) Texture-suppressed image (c) Superpixel segmentation (d) Estimated background(superpixels in black) (e) Superpixel saliency (f) Per-pixel saliency.

trast with the estimated background. Finally, the saliency of each pixel is assigned as a weighted linear combination of the saliency of its surrounding superpixels.

2.1. Texture Suppression

When human detect salient regions in natural scenes, extensively existed small scale textures, such as grass and foliage, do not catch attention at first glance. Some physiology experiments [15] have also shown that human vision system is more sensitive to patterns in middle frequencies than those in high frequencies. For example, we are attracted by the orange flower when seeing Fig. 1(a) at first glance, while ignore the small variations of leaves in the right-bottom region. However, the contributions of such little variations within textures may accumulate and influence the saliency computation in computational models.

We employ the structure extraction algorithm [16] to smooth out the local gradients in textures while preserve the global structures of the objects. The objective function in [16] is expressed as:

$$\arg\min_{S} \sum_{p} (S_p - I_p)^2 + \lambda \cdot \left(\frac{\mathcal{D}_x(p)}{\mathcal{L}_x(p) + \varepsilon} + \frac{\mathcal{D}_y(p)}{\mathcal{L}_y(p) + \varepsilon}\right), \quad (1)$$

where I is the input image, p indexes 2D pixels, S is the resulting structure image, $\mathcal{D}_x(p)$ and $\mathcal{D}_y(p)$ are sum of absolute spatial difference in the x and y directions weighted by a gaussian function within a window for pixel p, $\mathcal{L}_p(x)$ and $\mathcal{L}_p(y)$ are modulo of sum of directional spatial difference in the x and y directions weighted by a gaussian function within a window for pixel p. The first term $(S_p - I_p)^2$ is the smoothing term which ensures the smoothed image not deviate too much from the input. The second term is the regularizer named as *relative total variation* which enhances the contrast between texture and structure. Different values of the parameter λ in Eq. 1 produce images with various smoothness. Larger the λ is, smoother the image is. As shown in Fig. 1(b) ($\lambda = 0.05$), both background and foreground become more homogeneous after suppressing the textures, while the dissimilarities between foreground and background are preserved.

The texture-suppressed image is then segmented into non-overlapping superpixels, which have homogeneous color or texture. One property of superpixel is preserving object boundaries: all pixels in a superpixel mostly belong to the same object or stuff. We use *Simple Linear Iterative Clustering* (SLIC) [17], which adheres to boundaries well, to get the superpixels as shown in Fig. 1(c).

2.2. Saliency Estimation via Background Contrast

We estimate the background based on the background prior that the image boundaries are mostly background and salient objects rarely touch image boundaries. The soundness of such prior can be found in [11].

For each image, we define the estimated boundary, $\hat{B}$, as the superpixels involving any pixel whose closest distance to image boundaries is within n pixels. That is,

$$\hat{B} = \{S_i | min\{x, y, |W - x|, |H - y|\} \leq n, \exists I_p \in S_i\}, \quad (2)$$

where (x, y) denotes the position of a pixel I_p, S_i is a superpixel, W and H are the width and height of the input image, respectively. We choose $n = 10$ as a typical estimation. An exemplar of estimated background is shown in Fig. 1(d).

The *background contrast* of each superpixel is defined to be the summation of its kNN color distances with the superpixels in the estimated boundary $\hat{B}$. We use the CIELab color space and the distance is measured in Euclidean space.

Let the position and color of S_i be $\mathbf{p}_i$ and $\mathbf{c}_i$ which are the average of those pixels belong to the superpixel, respectively. The distances between superpixel S_i and every superpixel S_j in the estimated background $\hat{B}$ are denoted as

$$D_i = \{\|\mathbf{c}_i - \mathbf{c}_j\|^2 \mid \forall S_j \in \hat{B}\}. \quad (3)$$

Then we calculate a permutation (reordering) of the distance set D_i as

$$\tilde{D}_i = < d_{i1}, d_{i2}, \cdots, d_{iM} >, d_{i1} \leq d_{i2} \leq \cdots \leq d_{iM}. \quad (4)$$

where M is the number of superpixels in $\hat{B}$. We define the saliency of each superpixel as the summation of its kNN color distances with the estimated background superpixels, namely the background contrast, that is

$$Sa(S_i) = \sum_{j=1}^{k} d_{ij}. \quad (5)$$

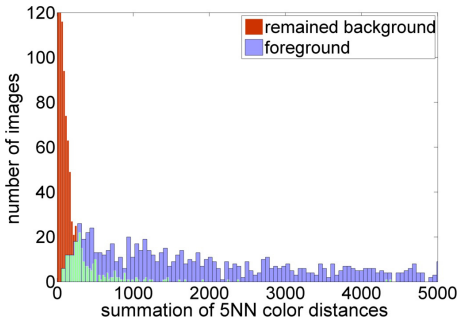

Fig. 2. Background contrast of the foreground and the remained background in MSRA-1000 dataset [7]. The red histogram and the blue histogram show the distributions of background contrast of the remained background and foreground, respectively. The overlapped bars are highlighted in green. Horizontal axis stands for background contrast values.

We denote the superpixels in foreground area of ground truth labeling as F, and the superpixels in remained background (namely the background area in ground truth labeling excluding the estimated background) as B. We verify our saliency definition by comparing the background contrast of both F and B. We calculate the average background contrast of superpixels in F and B of each image in MSRA-1000 dataset [7]. The histogram of background contrast of the remained background and that of the foreground are illustrated in Fig. 2. The parameters used in the experiment are $\lambda = 0.05$, $N = 100$, $k = 5$. N denotes the number of superpixels which does not influence the final results much, so we fix N to be 100 in subsequent experiments as well.

As shown in the red histogram, average background contrast of remained background B are mostly near zero, while the average background contrast of foreground F shown in blue histograms are larger. Some conclusions can be obtained that: (1) In a texture-suppressed image, the estimated background based on background prior is able to represent most background variations in the image. (2) Superpixels in F and B are approximately separable using background contrast.

Therefore the saliency defined in Eq. 5 is expected to assign large values to the superpixels in foreground and small values to the ones in background. Then we can separate the foreground and background to some extent.

Finally, a per-pixel saliency map is obtained by exploiting the fact that pixels with similar position and color shall have similar saliency values. We define the saliency of each pixel as a weighted linear combination of the saliency of its surrounding superpixels. The initial saliency value of each pixel is assigned as that of its superpixel. Then the refined saliency map is computed by choosing a gaussian weight that involves position and color information similar as in [10].

3. EXPERIMENTS

We evaluate our approach and compare it with several state-of-the-art methods on the publicly available MSRA-1000 dataset of 1000 images provided by Achanta *et al.* [7], which provides human labeled object segmentation masks.

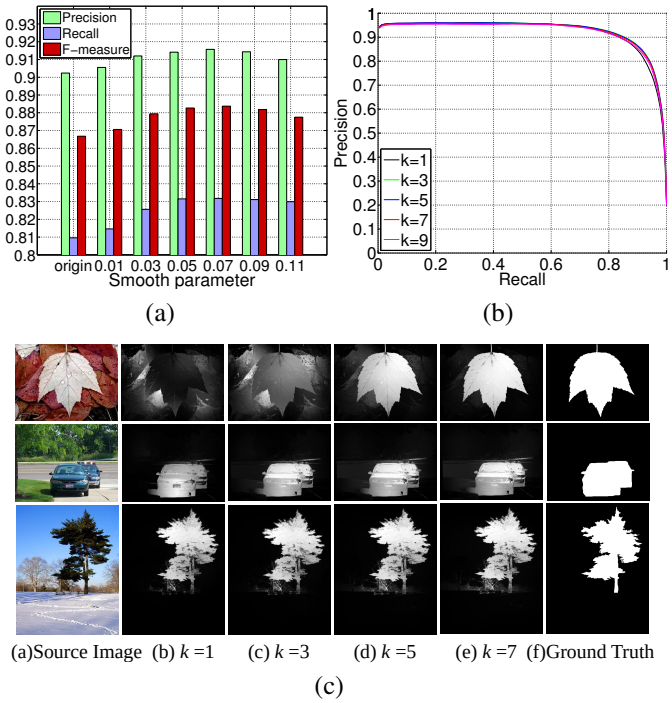

Fig. 3. Experimental results with different parameters. (a) Precision,recall and F-measure of different texture smooth parameters λ in Eq. 1 with adaptive threshold, $N = 100, k = 5$. (b) PR curves with different k in Eq. 5, $N = 100, \lambda = 0.05$. (The curves are nearly overlapped.) (c) Exemplars of saliency map with different k.

We adopt the same performance evaluation methods as [7] and [10]. In the first evaluation, a precision-recall (PR) curve of a saliency map is obtained by varying the threshold from 0 to 255. Precision measures the percentage of salient pixels correctly assigned, while recall is the fraction of detected salient pixels with respect to salient pixels in ground truth. The average precision-recall curve is generated by combining the results of all the 1000 test images.

In the second evaluation, we compare the performance when applying the adaptive threshold T proposed by [7], defined as twice the mean saliency of the image in Eq. 6, where W and H are the width and height of the input image, respectively. In addition to precision and recall, we compute their weighted harmonic mean measure of F-measure defined by Eq. 7. We set $\beta^2 = 0.3$ to weigh precision more than recall as suggested in [7, 9].

$$T = \frac{2}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} Sa(x, y) \tag{6}$$

$$F_\beta = \frac{(1 + \beta^2) \cdot Precision \cdot Recall}{\beta^2 \cdot Precision + Recall} \tag{7}$$

3.1. Performance of different parameters

We investigate the influence of parameters in our algorithm. The precision, recall and F-measure using adaptive threshold with different texture smoothness λ in Eq. 1 are shown in Fig. 3(a). We can see that the system with $\lambda \in [0.05, 0.09]$ performs fairly well. Compared with original images, texture-suppressed images yield better results. Texture suppression makes the background and foreground more homogeneous which enables the saliency computation to be more robust to the small variations in both foreground and background.

The PR curves with different k in Eq. 5 are shown in Fig. 3(b). The performances with different k do not change much, while the results with $k = 3, 5, 7$ are a little better than those with $k = 1$ and $k = 9$. The proper k (e.g. $k = 5$) can deal with corruptions in background estimation, thus gets more robust results in the case that foreground patterns appear in image boundaries. Please note that [11] initializes the saliency of image boundaries by using a saliency detection algorithm [18] to boundary patches to alleviate such problem. Some exemplars are shown in Fig. 3(c). We can see that proper k performs more robustly when dealing with background prior failure. We choose $k = 5$ for subsequent experiments.

3.2. Evaluation of our approach on MSRA-1000 dataset

We compare our method(TB) with some other approaches: the spectral residual approach(SR)[8], Zhai's method (LC)[13], frequency-tuned approach(FT)[7], histogram based contrast(HC) and region based contrast(RC)[9], geodesic saliency(GS)[11](we use GS_SP in following comparisons), saliency filters(SF)[10] and the low rank matrix recovery approach(LR)[14]. Each method outputs a full resolution saliency map that is normalized to range $[0, 255]$. The PR curves of all the approaches on the MSRA-1000 dataset[7], as well as comparisons of precision, recall and F-measure with adaptive threshold are shown in Fig. 4(a) and (b), respectively. Some exemplars of these methods are illustrated in Fig. 4(c).

We conclude from the comparisons that: (1) Our algorithm outperforms the previous methods. Especially, it performs robustly when dealing with textured and cluttered images as shown in Fig. 4(c). (2)Background contrast with texture suppression is an effective measurement to distinguish foreground from background, as well as decrease the saliency values of background.

4. CONCLUSIONS

We propose a salient region detection algorithm by texture-suppressed background contrast and improve the state-of-the-art approaches. Inspired biologically by the fact that human vision system is not sensitive to patterns in high frequencies, we suppress the small variations in texture regions, which proves to be a good way of exploiting the homogeneous nature of the background. We define the saliency as the summation of k minimum distances based on background contrast, which is robust and achieves better overall performance. In our future work, we will investigate discriminative descriptions of foreground and background and obtain more accurate estimation of the background.

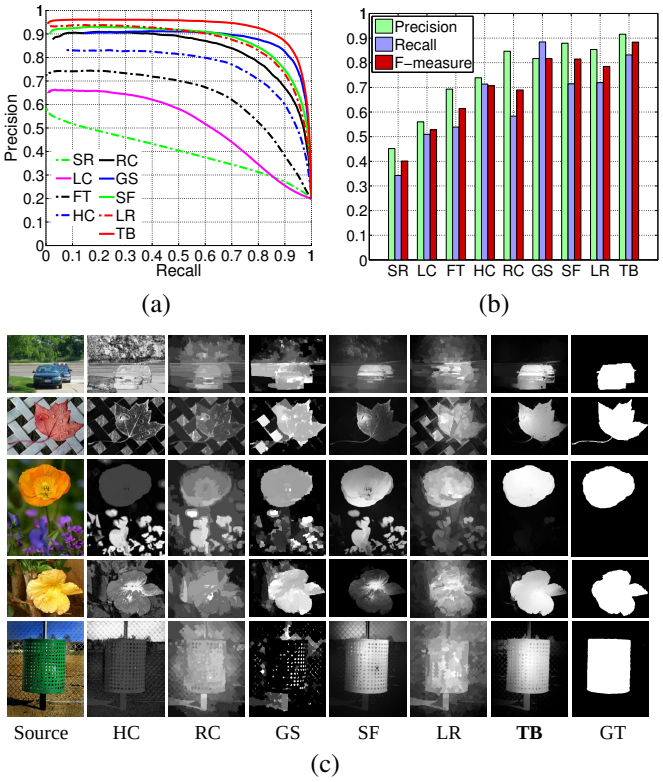

Fig. 4. Results on the MSRA-1000 dataset[7]. (a) PR curve compared to other methods. (b) Comparison to other methods of precision, recall and F-measure when applying adaptive threshold. (c) Visual comparison of previous approaches to our method (TB) and ground truth (GT). Due to space limit, only the results from five other methods that give good PR curves in (a) are presented.

5. REFERENCES

[1] Bogdan Alexe, Thomas Deselaers, and Vittorio Ferrari, "What is an object?," in *CVPR*, 2010, pp. 73–80.

[2] Michael Donoser, Martin Urschler, Martin Hirzer, and Horst Bischof, "Saliency driven total variation segmentation," in *ICCV*, 2009, pp. 817–824.

[3] Vijay Mahadevan and Nuno Vasconcelos, "Saliency-based discriminant tracking," in *CVPR*, 2009, pp. 1007–1013.

[4] Laurent Itti, Christof Koch, and Ernst Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, 1998.

[5] Dashan Gao, Vijay Mahadevan, and Nuno Vasconcelos, "The discriminant center-surround hypothesis for bottom-up saliency," in *NIPS*, 2007.

[6] Lijuan Duan, Chunpeng Wu, Jun Miao, Laiyun Qing, and Yu Fu, "Visual saliency detection by spatially weighted dissimilarity," in *CVPR*, 2011, pp. 473–480.

[7] Radhakrishna Achanta, Sheila S. Hemami, Francisco J. Estrada, and Sabine Süsstrunk, "Frequency-tuned salient region detection," in *CVPR*, 2009, pp. 1597–1604.

[8] Xiaodi Hou and Liqing Zhang, "Saliency detection: A spectral residual approach," in *CVPR*, 2007.

[9] Ming-Ming Cheng, Guo-Xin Zhang, Niloy J. Mitra, Xiaolei Huang, and Shi-Min Hu, "Global contrast based salient region detection," in *CVPR*, 2011, pp. 409–416.

[10] Federico Perazzi, Philipp Krähenbühl, Yael Pritch, and Alexander Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *CVPR*, 2012, pp. 733–740.

[11] Yichen Wei, Fang Wen, Wangjiang Zhu, and Jian Sun, "Geodesic saliency using background priors," in *ECCV (3)*, 2012, pp. 29–42.

[12] Jie Feng, Yichen Wei, Litian Tao, Chao Zhang, and Jian Sun, "Salient object detection by composition," in *ICCV*, 2011, pp. 1028–1035.

[13] Yun Zhai and Mubarak Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *ACM Multimedia*, 2006, pp. 815–824.

[14] Xiaohui Shen and Ying Wu, "A unified approach to salient object detection via low rank matrix recovery," in *CVPR*, 2012, pp. 853–860.

[15] F. W. Campbell and J. G. Robson, "Application of fourier analysis to the visibility of gratings.," *J Physiol*, vol. 197, no. 3, pp. 551–566, Aug. 1968.

[16] Li Xu, Qiong Yan, Yang Xia, and Jiaya Jia, "Structure extraction from texture via relative total variation," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 139, 2012.

[17] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurélien Lucchi, Pascal Fua, and Sabine Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, 2012.

[18] Stas Goferman, Lihi Zelnik-Manor, and Ayellet Tal, "Context-aware saliency detection," in *CVPR*, 2010, pp. 2376–2383.