Saliency detection using regional histograms

Zhi Liu,[1,2,]* Olivier Le Meur,[2,3] Shuhua Luo,[1] and Liquan Shen[1]

[1]*School of Communication and Information Engineering, Shanghai University, Shanghai 200072, China*

[2]*IRISA, Campus Universitaire de Beaulieu, Rennes 35042, France*

[3]*University of Rennes 1, Campus Universitaire de Beaulieu, Rennes 35042, France*

Corresponding author: liuzhisjtu@163.com

We propose an efficient regional histogram (RH)-based computation model for saliency detection in natural images. First, the global histogram is constructed by performing an adaptive color quantization on the original image. Then multiple RHs are built on the basis of the region segmentation result, and the color–spatial similarity between each pixel and each RH is calculated accordingly. Two efficient measures, distinctiveness and compactness of each RH, are evaluated based on the color difference with the global histogram and the color distribution over the whole image, respectively. Finally, the pixel-level saliency map is generated by integrating the color–spatial similarity measures with the distinctiveness and compactness measures. Experimental results on a dataset containing 1000 test images with ground truths demonstrate that the proposed saliency model consistently outperforms state-of-the-art saliency models.   © 2013 Optical Society of America

   OCIS codes: 100.2960, 100.2000, 330.4060.

Saliency detection plays an important role in a variety of applications including salient object detection, salient object segmentation, and content-aware image retargeting, etc. A number of saliency models implement the center-surround scheme, which has a clear interpretation of visual attention mechanism, with variations of features and surround selections [1–4]. Besides, there are various formulations for measuring saliency based on information theory [5], graph theory [6], supervised learning [7], and frequency domain analysis [8], etc. Recently, different forms of global information on the basis of region segmentation are efficiently incorporated into saliency models to generate saliency maps (SMs) with well-defined boundaries [9–11]. However, the existing saliency models are still insufficient to efficiently suppress irrelevant background regions and completely highlight salient object regions in complicated images with cluttered background, highly textured regions, and low contrast between objects and background.

In this Letter, we propose a novel saliency model based on regional histograms (RHs) with the aim to improve the saliency detection performance. Specially, we first propose for RHs two efficient measures, distinctiveness and compactness, which systematically integrate the factors of color contrast with the global histogram, color distribution over the whole image, and spatial locations of different RHs. Then we exploit the distinctiveness and compactness measures of RHs to generate pixel-level SMs with higher subjective visual quality. Objective evaluations also demonstrate that the proposed saliency model achieves a consistently higher saliency detection performance compared to state-of-the-art saliency models.

The original color image in the RGB color space is first transformed into the Lab color space, in which the luminance channel and the two chrominance channels are separated. Each channel is then uniformly quantized into q bins, and the global histogram H_0 with $q \times q \times q$ bins is calculated using all pixels in the image. The parameter q is set to a moderate value, 16, which is generally sufficient for color quantization of natural images. The quantized color of each bin, $\mathbf{c}_k (k = 1, 2, \ldots, q^3)$ is calculated as

the mean color of those pixels that fall into the kth bin. In order to moderately reduce the number of colors in the global histogram, all bins in H_0 are first sorted in the descending order, and the first m bins, which cover just more than α percentage of pixels, are selected as the representative colors. Then each of remaining bins in H_0 is merged into one of the selected m bins with the smallest difference between their quantized colors, and the quantized color of the selected bin is updated accordingly. Finally, a color quantization table with m entries is generated for the image. The parameter α controls the number of quantized colors in the global histogram, and is set to 95, which is sufficient to preserve the representative colors well. For the example image in Fig. 1(a), its global histogram is shown in the top-left corner (with the label "0") of Fig. 2(a), in which each bar of the histogram is represented using the quantized color. For a clear display, only a part of bins with higher values are shown.

As an efficient local representation of the image, multiple RHs are built on the basis of region segmentation result, which is generated using the mean shift algorithm [12]. We only adjust the parameter β, which is the minimum allowable region area used in the mean shift algorithm, to simply control the degree between oversegmentation and undersegmentation. In our implementation, we set β to 2 percentage of image area, for a moderate region segmentation result on most images. For the example image in Fig. 1(a), the region segmentation result is shown in the center of Fig. 2(a), in which each segmented region is represented using the region's mean color.

Using the color quantization table of the image, each pixel is assigned with one of the m quantized
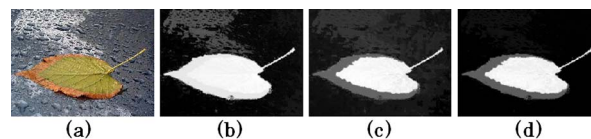

Fig. 1. (Color online) (a) Original image, (b) pixel-level distinctiveness map, (c) pixel-level compactness map, and (d) pixel-level saliency map.
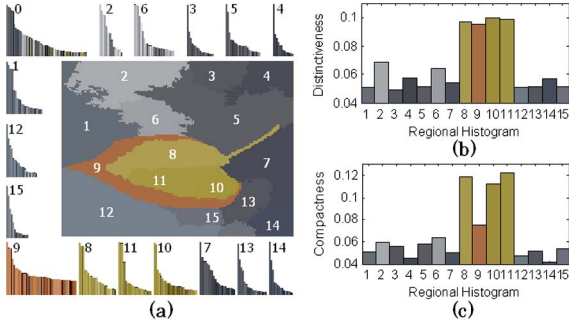
Fig. 2. (Color online) (a) Illustration of global histogram and multiple RHs, (b) distinctiveness measures, and (c) compactness measures of RHs.

colors. For each segmented region $R_i(i = 1, \ldots n)$, its RH $H_i(i = 1, \ldots n)$ is then calculated based on the quantized colors of all pixels in R_i. The spatial position μ_i of each RH H_i is defined as the geometric center of R_i. The color–spatial similarity between each pixel p and each RH H_i is then defined as

$$C(p, H_i) = H_i(b_p) \cdot D_H(p, H_i), \qquad (1)$$

where b_p denotes the bin number for the quantized color of p in H_i, and $D_H(p, H_i)$ is defined as

$$D_H(p, H_i) = 1 - \frac{\|\mathbf{x}_p - \boldsymbol{\mu}_i\|}{d}, \qquad (2)$$

where d denotes the diagonal length of the image, and $\mathbf{x}_p$ denotes the spatial coordinates of p. The value of $D_H(p, H_i)$ falls into the range of [0, 1], and $C(p, H_i)$ achieves a higher value when the probability of the pixel's quantized color is higher in a RH, which is also spatially near to the pixel.

As shown in Fig. 2(a), the original image is partitioned into 15 regions, and thus can be represented using 15 RHs located at different spatial positions. We can observe from Fig. 2(a) that the color distributions of these RHs (No. 8–11) located at the salient object are different from other RHs located at the background.

We observed from a variety of natural images that the colors of salient objects are distinctive in the whole image, and salient objects are generally surrounded by background regions. Therefore, the RHs that mainly contain the colors of salient objects are usually distinctive from the global histogram, and the colors of salient objects have a more compact distribution over the whole image than background colors. Based on the above two aspects, two terms, i.e., distinctiveness and compactness of RHs, are proposed to measure the saliency of RHs.

Specifically, the distinctiveness of each RH H_i is evaluated based on its difference with the global histogram H_0 as follows

$$RD(H_i) = \sum_{j=1}^{m} \left[H_i(j) \sum_{k=1}^{m} \|\mathbf{c}_j - \mathbf{c}_k\| \cdot H_0(k) \right], \qquad (3)$$

where $H_i(j)$ is the probability of the jth bin in H_i. Using Eq. (3), the difference between H_i and H_0 is evaluated based on the differences between different quantized colors weighted by their occurrence probabilities.

Different colors in each RH H_i have different spatial distributions over the whole image. Therefore, the spread degree of the colors in H_i is evaluated based on the distribution of pixels with similar colors over the whole image, and is defined as

$$RS(H_i) = \frac{\sum_{p \in I} C(p, H_i) \cdot D_O(p)}{\sum_{p \in I} C(p, H_i)}, \qquad (4)$$

where I denotes the whole image, and $D_O(p)$ denotes the Euclidean spatial distance from the pixel p to the center position of the image. The compactness of H_i is then defined as the reciprocal of Eq. (4), i.e.,

$$RC(H_i) = \frac{1}{RS(H_i)}. \qquad (5)$$

For the example image in Fig. 1(a), the above two measures, distinctiveness and compactness, are calculated for all RHs, and their normalized values ($\sum_{i=1}^{n} RD(H_i) = 1$ and $\sum_{i=1}^{n} RC(H_i) = 1$) are shown in Figs. 2(b) and 2(c), respectively. We can see that the RHs (No. 8–11) achieve higher values on both distinctiveness and compactness, while the two measures of other RHs are lowered.

Based on distinctiveness and compactness measures of RHs, the pixel-level distinctiveness map (DM) and compactness map (CM) are defined as follows

$$\mathrm{DM}(p) = \sum_{i=1}^{n} C(p, H_i) \cdot RD(H_i), \qquad (6)$$

$$\mathrm{CM}(p) = \sum_{i=1}^{n} C(p, H_i) \cdot RC(H_i). \qquad (7)$$

It can be seen from Eqs. (6) and (7) that for measuring the distinctiveness and compactness of each pixel p, the contribution of each RH H_i is determined by the corresponding color–spatial similarity $C(p, H_i)$. The distinctiveness/compactness of each pixel is the sum of distinctiveness/compactness measures of all RHs weighted by the corresponding color–spatial similarity measures.

Finally, by performing a pixel-wise multiplication operation between the DM and the CM, the pixel-level SM is generated as follows:

$$\mathrm{SM}(p) = DM(p) \cdot CM(p). \qquad (8)$$

The pixel-level DM, CM, and SM for Fig. 1(a) are respectively shown in Figs. 1(b)–1(d), which are normalized into the same range of [0, 255] for display. Compared with Figs. 1(b) and 1(c), the background regions are more efficiently suppressed in Fig. 1(d). We found that for most images, the DM and the CM can complement each other to generate a high-quality SM, which can highlight salient objects and suppress background regions more effectively.

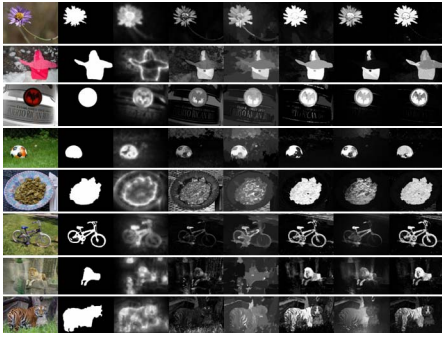We performed experiments on an image dataset [2] with manually segmented ground truths for salient

Fig. 3.   (Color online) Some examples of saliency detection. From left to right: original images, ground truths, and saliency maps generated using CA, FT, RC, KD, SF, and RH, respectively.

Fig. 4.   (Color online) Precision–recall curves of different saliency models.

objects in 1000 images, which were selected from Microsoft Research Asia Salient Object Database (MSRA SOD) Image Set B [7]. We compared our RH-based saliency model with five state-of-the-art saliency models including a context-aware (CA) model [3], frequency-tuned (FT) model [2], region contrast (RC)-based model [9], kernel density estimation (KD)-based model [10], and contrast-based saliency filter (SF) [11]. We used executables, source codes with default parameter settings or SMs provided by the authors for the five saliency models. For a fair comparison, all SMs are normalized into the same range of [0, 255] with the full resolution of original images. Some SMs generated using these saliency models are shown in Fig. 3. Compared with other saliency models, we can observe from Fig. 3 that salient object regions can be more completely highlighted and background regions can be more effectively suppressed in the SMs generated using our saliency model RH.

In order to objectively evaluate the saliency detection performance of different saliency models, we adopted the commonly used performance measures, i.e., precision and recall. The thresholding operations using a series of fixed integers from 0 to 255 are first performed on each SM to obtain 256 binary salient object masks, and a set of precision/recall measures are calculated using the binary ground truth as the reference mask. Then for each saliency model, at each threshold, the precision/recall measures of all 1000 SMs are averaged, and as shown in Fig. 4, the precision-recall curve for each saliency model plots the 256 average precision measures against the 256 average recall measures. We can see from Fig. 4 that the precision–recall curve of RH is higher than the other precision–recall curves, and thus can objectively demonstrate that RH outperforms the other saliency models on saliency detection performance.

In conclusion, we have presented a novel saliency model by evaluating the distinctiveness and compactness measures of RHs in this Letter, and both subjective and objective evaluations demonstrate that our saliency model consistently outperforms the existing state-of-the-art saliency models.
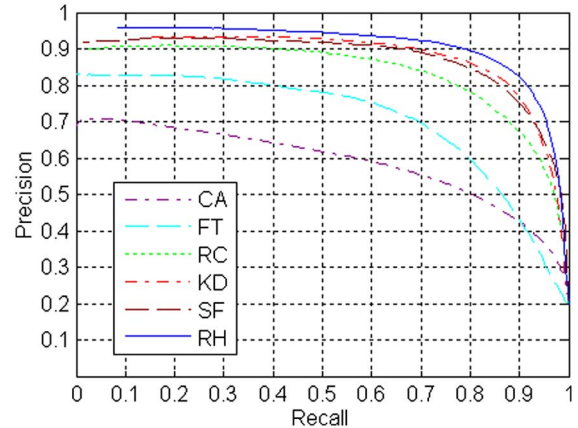
References

1. L. Itti, C. Koch, and E. Niebur, IEEE Trans. Pattern Anal. Mach. Intell. **20**, 1254 (1998).
2. R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2009), pp. 1597–1604.
3. S. Goferman, L. Zelnik-Manor, and A. Tal, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2010), pp. 2376–2383.
4. W. Kim and C. Kim, Opt. Lett. **37**, 1550 (2012).
5. L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell, J. Vis. **8**(7), 32 (2008).
6. J. Harel, C. Koch, and P. Perona, in *Proceedings of NIPS* (MIT, 2006) pp. 545–552.
7. T. Liu, J. Sun, N. Zheng, X. Tang, and H. Y. Shum, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2007), paper 4270072.
8. X. Hou and L. Zhang, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2007), paper 4270292.
9. M. Cheng, G. Zhang, N. Mitra, X. Huang, and S. Hu, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2011), pp. 409–416.
10. Z. Liu, R. Shi, L. Shen, Y. Xue, K. N. Ngan, and Z. Zhang, IEEE Trans. Multimedia **14**, 1275 (2012).
11. F. Perazzil, P. Krähenbül, Y. Pritch, and A. Hornung, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2012), pp. 733–740.
12. D. Comaniciu and P. Meer, IEEE Trans. Pattern Anal. Mach. Intell. **24**, 603 (2002).