SUPERPIXEL-BASED SALIENCY DETECTION

Zhi Liu[1,2], Olivier Le Meur[2,3], Shuhua Luo[1]

[1]School of Communication and Information Engineering, Shanghai University, Shanghai, China
[2]IRISA, Campus Universitaire de Beaulieu, Rennes, France
[3]University of Rennes 1, Campus Universitaire de Beaulieu, Rennes, France

ABSTRACT

In this paper, we propose an effective superpixel-based saliency model. First, the original image is simplified by performing superpixel segmentation and adaptive color quantization. On the basis of superpixel representation, inter-superpixel similarity measures are then calculated based on difference of histograms and spatial distance between each pair of superpixels. For each superpixel, its global contrast measure and spatial sparsity measure are evaluated, and refined with the integration of inter-superpixel similarity measures to finally generate the superpixel-level saliency map. Experimental results on a dataset containing 1,000 test images with ground truths demonstrate that the proposed saliency model outperforms state-of-the-art saliency models.

1. INTRODUCTION

Visual saliency is the perceptual quality that makes some objects in the scene stand out from their surrounding regions and thus capture human visual attention. Visual attention mechanism allows us to effortlessly identify visually salient objects even in a complex scene. With the goal both to achieve a comparable saliency detection performance as our visual system and to facilitate different saliency-based applications such as salient object detection/segmentation and content-aware image/video retargeting, a number of saliency models have been proposed in the past decades.

Itti et al. were the first to propose a biologically plausible saliency model [1], which first computes feature maps of luminance, color and orientation using a center-surround operator across different scales, and then performs normalization and summation to generate the saliency map. Since then, the center-surround scheme has been implemented using more features such as Kullback-Leibler divergence between histograms of filter responses [2], oriented subband decomposition based energy [3], and local regression kernel based self-resemblance [4]. Specially, in the frequency-tuned saliency model [5], the whole image after blurring is used as the surrounding region.

Besides, various formulations based on information theory, graph theory, supervised learning, and frequency domain analysis have been proposed for measuring saliency. For example, saliency is measured based on the rarity represented using self-information of local image features in [6], random walks on the weighted graph constructed from the image in [7], supervised learning for integrating a set of features under the framework of conditional random field in [8], and the spectral residual of the amplitude spectrum of Fourier transform in [9].

Recently, the global information of image is incorporated into saliency models with different forms. In the context-aware saliency model [10], the global uniqueness of color features and some visual organization rules are combined with the local center-surround difference. In [11], both global color distribution and orientation distribution are fully utilized to selectively generate the saliency map. Furthermore, in some recent saliency models [12-15], the image is partitioned into regions using either image segmentation or pixel clustering, and the global information is efficiently incorporated at region level using global contrast and spatially weighted regional contrast measures [12], color compactness measure [13] and statistical models such as Gaussian model [14] and kernel density estimation based nonparametric model [15].

However, the state-of-the-art saliency models are still insufficient to effectively highlight salient object regions completely with well-defined boundaries and effectively suppress background especially for some complicated images. In this paper, we propose a simple yet effective superpixel-based saliency model, in which superpixels are used as the basic primitives for saliency measuring. Based on the simplified image using superpixel representation and adaptive color quantization, we propose inter-superpixel similarity measure, global contrast measure and spatial sparsity measure to derive superpixel-level saliency map, which better highlights salient objects and suppresses background regions more effectively. Objective evaluations also demonstrate that the proposed saliency model achieves a consistently higher saliency detection performance than state-of-the-art saliency models.

The rest of this paper is organized as follows. Section 2 describes the proposed superpixel-based saliency model in detail. Experimental results are presented in Section 3, and conclusions are given in Section 4.
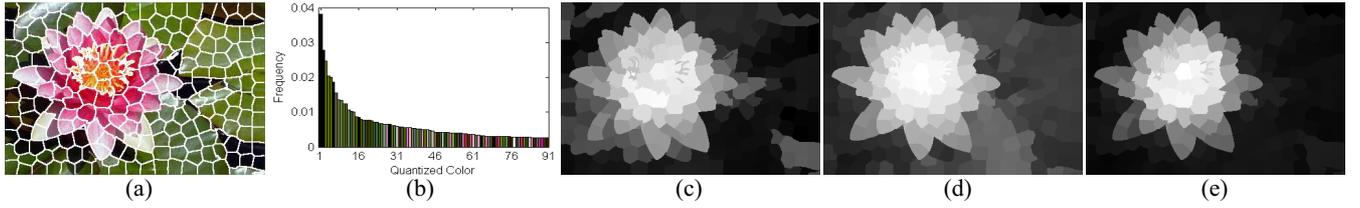
Fig. 1. (a) Superpixel segmentation result, (b) image histogram, (c) the refined global contrast map,(d) the refined spatial sparsity map, and (e) final superpixel-level saliency map.

2. SUPERPIXEL-BASED SALIENCY MODEL

The proposed saliency model consists of the following three stages. First, superpixel segmentation and adaptive color quantization are used to simplify the input image. Then, inter-superpixel similarity between each pair of superpixels is calculated based on their histogram difference and spatial distance. Finally, two effective measures of global contrast and spatial sparsity are evaluated for each superpixel, and finally integrated with the inter-superpixel similarity measures to generate the superpixel-level saliency map.

2.1. Image simplification

We perform two simplification operations, i.e., superpixel segmentation and adaptive color quantization, to represent the image using a reduced number of superpixels and representative colors. The original image is first transformed into the *Lab* color space, in which luminance channel and two chrominance channels are separated.

We choose the recently proposed superpixel algorithm based on simple linear iterative clustering [16] to partition the original image into a number of superpixels, which usually have a regular and compact shape with better boundary adherence. The parameter of the starting size of superpixels is set to $\sqrt{N/200}$, where N denotes the number of pixels in the image, and the number of the generated superpixels is approximately 200, which is generally sufficient to preserve different boundaries well. An example of superpixel segmentation on an image is shown in Fig. 1(a), in which the boundaries of different superpixels are delineated using white lines.

For adaptive color quantization, each color channel is first quantized into q bins to generate the image histogram H_0 with $q \times q \times q$ bins, and the quantized color of each bin, $\mathbf{qc}_k$ is calculated as the mean color of those pixels that fall into the kth bin. Then the first m high-probability bins that cover just more than $\alpha \cdot N$ pixels, are selected as the representative colors. Finally, each of the remaining $q^3 - m$ bins is merged into one of the selected m bins with the smallest difference between their quantized colors, and the quantized colors of all bins are updated to generate a color quantization table Q with m entries. The parameters q and α are set to 16 and 0.95, respectively, for a moderate color quantization of natural images. For the example in Fig. 1(a), Fig. 1(b) shows a part of H_0, in which each bin is represented using its quantized color.

2.2. Inter-superpixel similarity

Using all pixels in each superpixel $SP_i (i = 1,...,n)$, the superpixel-level histogram H_i is calculated and simplified based on Q, and then normalized to have $\sum_{k=1}^{m} H_i(k) = 1$. The inter-superpixel similarity between each pair of superpixels, SP_i and SP_j, is defined as

$$Sim(i,j) = Sim_c(i,j) \cdot Sim_d(i,j) \tag{1}$$

where the color similarity $Sim_c(i,j)$ is defined as the sum of histogram intersection between H_i and H_j,

$$Sim_c(i,j) = \sum_{k=1}^{m} \min\{H_i(k), H_j(k)\} \tag{2}$$

and the spatial similarity $Sim_d(i,j)$ is defined as

$$Sim_d(i,j) = 1 - \frac{\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|}{d} \tag{3}$$

where d denotes the diagonal length of the image, and $\boldsymbol{\mu}_i$ denotes the spatial center position of SP_i.

2.3. Superpixel saliency

We observed from a variety of natural images that salient object superpixels usually show noticeable color contrast with background superpixels, and the spatial distribution of salient object superpixels is sparser than background superpixels, which usually scatter over the whole image. Based on these two observations, the global contrast and spatial sparsity of superpixels are evaluated for measuring the saliency of superpixels.

The global contrast of SP_i is measured using the weighted color differences with all the other superpixels,

$$GC(i) = \sum_{j=1}^{n} W(i,j) \cdot \|\mathbf{mc}_i - \mathbf{mc}_j\| \tag{4}$$

where $\mathbf{mc}_i$ denotes the mean color of SP_i. The weight $W(i,j)$ is defined by considering the factors of superpixel area and spatial similarity as follows:

$$W(i,j) = |SP_j| \cdot Sim_d(i,j) \tag{5}$$

where |.| denotes the number of pixels in the superpixel. The normalized global contrast measure for SP_i is calculated as follows:

$$NGC(i) = \frac{GC(i) - GC_{\min}}{GC_{\max} - GC_{\min}} \qquad (6)$$

where $GC_{\max} / GC_{\min}$ is the maximum/minimum in the global contrast measures of all superpixels.

For each superpixel SP_i, the spatial spread of its color distribution is defined as

$$SS(i) = \frac{\sum_{j=1}^{n} Sim(i,j) \cdot D(j)}{\sum_{j=1}^{n} Sim(i,j)} \qquad (7)$$

where $D(j)$ denotes the Euclidean spatial distance from the center position of SP_j to the image center position. Then an inverse normalization operation is performed on the spatial spread measures to obtain the normalized spatial sparsity measure for each superpixel as follows:

$$NSS(i) = \frac{SS(i) - SS_{\max}}{SS_{\min} - SS_{\max}} \qquad (8)$$

where $SS_{\max} / SS_{\min}$ is the maximum/minimum in the spatial spread measures of all superpixels.

Based on a reasonable assumption that superpixels with higher similarity should be evaluated to have more similar saliency values, the inter-superpixel similarity measures are then exploited to refine the normalized global contrast/ spatial sparsity measure as follows:

$$RGC(i) = \frac{\sum_{j=1}^{n} Sim(i,j) \cdot NGC(j)}{\sum_{j=1}^{n} Sim(i,j)} \qquad (9)$$

$$RSS(i) = \frac{\sum_{j=1}^{n} Sim(i,j) \cdot NSS(j)}{\sum_{j=1}^{n} Sim(i,j)} \qquad (10)$$

For the example in Fig. 1(a), the refined global contrast measures (RGC) and the refined spatial sparsity (RSS) measures for all superpixels are respectively represented using Fig. 1(c) and (d), in which brighter superpixels indicate higher values of these two measures, and we can observe that salient object superpixels are highlighted and background superpixels are suppressed to some degree.

Finally, by performing a superpixel-wise multiplication operation between RGC and RSS, the saliency measure for each superpixel SP_i is defined as

$$Sal(i) = RGC(i) \cdot RSS(i) \qquad (11)$$

Based on Fig. 1(c) and (d), the superpixel-level saliency map generated using Eq. (11) is shown in Fig. 1(e), which is normalized into the range of [0, 255] for display. Compared with Fig. 1(c) and (d), the background superpixels are suppressed more effectively in Fig. 1(e), and the complete salient object is highlighted with well-defined boundaries.

3. EXPERIMENTAL RESULTS

We performed experiments on an image dataset [5] with manually segmented ground truths for salient objects in 1,000 images (publicly available at http://ivrg.epfl.ch/supplementary_material/RK_CVPR09/),
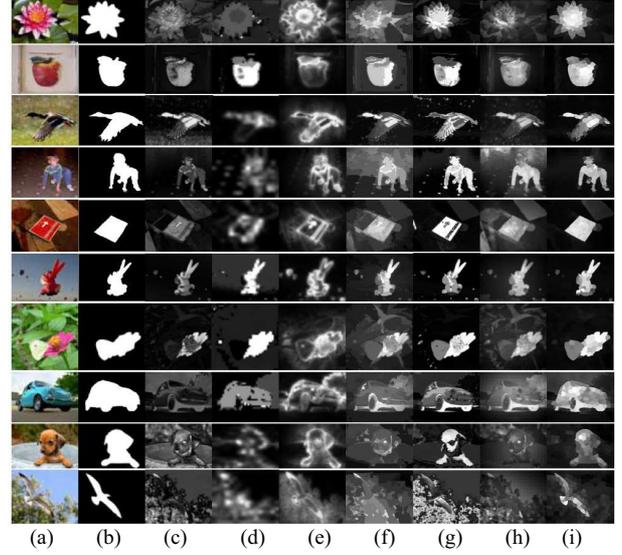

Fig. 2. (a) Original images, (b) ground truths, and saliency maps generated using saliency models including (c) FT, (d) CO, (e) CA, (f) RC, (g) KD, (h) OS and (i) SP, respectively.
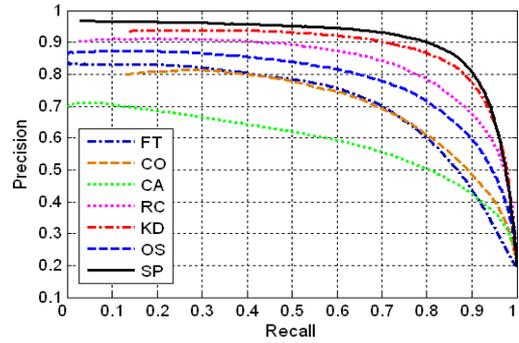

Fig. 3. Precision-recall curves of different saliency models.

which are selected from MSRA SOD Image Set B [8]. We compared our superpixel-based (SP) saliency model with six state-of-the-art saliency models including frequency-tuned (FT) [5], distribution of color and orientation (CO) [11], context-aware (CA) [10], region contrast (RC) [12], kernel density estimation (KD) [15], and over-segmentation (OS) [13] based saliency models. We used executables or source codes with default parameter settings provided by the authors for the other saliency models. For a fair comparison, all saliency maps are normalized into the same range of [0, 255] with the full resolution of original images.

For a subjective comparison, Fig. 2 shows some saliency maps generated using SP and the other six saliency models. Compared with other saliency models, we can observe from Fig. 2 that salient object regions can be better highlighted with well-defined boundaries, and background regions can be more effectively suppressed in the saliency maps generated using SP. For some complicated images such as the latter two examples, which show low contrast between salient object regions and some background regions, the saliency maps generated using SP still show a better visual quality than other saliency maps.
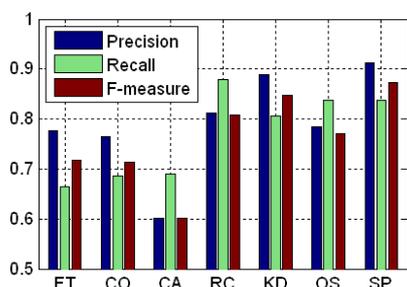
Fig. 4. Average value of precision, recall, F-measure achieved using different saliency models.

For an objective comparison, we adopted the commonly used performance measures, i.e., precision and recall, to evaluate the saliency detection performance of different saliency models. We used the fixed thresholding method in [5] to generate the precision-recall curve for each saliency model. As shown in Fig. 3, the precision-recall curve of SP is higher than all the other precision-recall curves, and thus can objectively demonstrate that SP outperforms the other saliency models on saliency detection performance.

In order to evaluate the applicability of saliency maps for salient object detection more explicitly, we performed adaptive thresholding operation on each saliency map using Otsu's method [17], which is simple yet effective, to obtain the binary salient object mask. We calculate the measures of precision and recall by comparing the obtained binary salient object mask with the corresponding binary ground truth, and then calculate F-measure, which is the harmonic mean of precision and recall, to evaluate the overall detection performance as follows:

$$F_\gamma = \frac{(1+\gamma)\cdot precision \cdot recall}{\gamma \cdot precision + recall} \quad (12)$$

where the coefficient γ is set to 0.3 as used in [5, 12]. The average precision, recall and F-measure of all saliency maps generated using each saliency model are shown in Fig. 4. We can see that SP achieves the highest value of F-measure, which demonstrates the overall better applicability of our saliency maps for salient object detection.

4. CONCLUSIONS

In this paper, we have presented an effective superpixel-based saliency model. By exploiting the proposed inter-superpixel similarity measure, global contrast measure and spatial sparsity measure, the superpixel-level saliency map generated using our saliency model can better highlight salient object and suppress background. Both subjective and objective experiments demonstrate that our saliency model outperforms six state-of-the-art saliency models.

5. ACKNOWLEDGMENT

6. REFERENCES

[1] L. Itti, C. Koch, and E. 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, Nov. 1998.

[2] D. Gao, V. Mahadevan, and N. Vasconcelos, "The discriminant center-surround hypothesis for bottom-up saliency," *Proc. NIPS*, pp. 497-504, Dec. 2007.

[3] O. Le Meur, and J.-C. Chevet, "Relevance of a feed-forward model of visual attention for goal-oriented and free-viewing tasks," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2801-2813, Nov. 2010.

[4] H. J. Seo, and P. Milanfar, "Static and space-time visual saliency detection by self-resemblance," *Journal of Vision*, vol. 9, no. 12, article 15, Nov. 2009.

[5] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," *Proc. IEEE CVPR*, pp. 1597-1604, Jun. 2009.

[6] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell, "Sun: A Bayesian framework for saliency using natural statistics," *Journal of Vision*, vol. 8, no. 7, article 32, Dec. 2008.

[7] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Proc. NIPS*, pp. 545-552, Dec. 2006.

[8] T. Liu, J. Sun, N. Zheng, X. Tang, and H. Y. Shum, "Learning to detect a salient object," *Proc. IEEE CVPR*, p. 4270072, Jun. 2007.

[9] X. Hou, and L. Zhang, "Saliency detection: a spectral residual approach," *Proc. IEEE CVPR*, p. 4270292, Jun. 2007.

[10] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *Proc. IEEE CVPR*, pp. 2376-2383, Jun. 2010.

[11] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Salient region detection by modeling distributions of color and orientation," *IEEE Trans. Multimedia*, vol. 11, no. 5, pp. 892-905, Aug. 2009.

[12] M. M. Cheng, G. X. Zhang, N. J. Mitra, X. Huang, and S. M. Hu, "Global contrast based salient region detection," *Proc. IEEE CVPR*, pp. 409-416, Jun. 2011.

[13] X. Zhang, Z. Ren, D. Rajan, and Y. Hu, "Salient object detection through over-segmentation," *Proc. IEEE ICME*, pp. 1033-1038, Jul. 2012.

[14] Z. Liu, Y. Xue, H. Yan, and Z. Zhang, "Efficient saliency detection based on Gaussian models," *IET Image Processing*, vol. 5, no. 2, pp. 122-131, Mar. 2011.

[15] Z. Liu, R. Shi, L. Shen, Y. Xue, K. N. Ngan, and Z. Zhang, "Unsupervised salient object segmentation based on kernel density estimation and two-phase graph cut," *IEEE Trans. Multimedia*, vol. 14, no. 4, pp. 1275-1289, Aug. 2012.

[16] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274-2282, Nov. 2012.

[17] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man Cybern.*, vol. 9, no. 1, pp. 62-66, Jan. 1979.