21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Corner-Surround Contrast for Saliency Detection

Quan Zhou, Nianyi Li, Yi Yang, Pan Chen and Wenyu Liu
Department of Electronics and Information Engineering,
Huazhong University of Science and Technology, Wuhan, P.R. China
{*qzhou.lhi, lnyhust, yiyang.cv, chenpan.male*}*@gmail.com, liuwy@mail.hust.edu.cn*

Abstract

Center-surround measurements are widely used for saliency detection but with some disadvantages: 1) Center-surround operation may cause inaccurate segmentation and even involve incorrect detection results; 2) In most situations, only using center-surround feature is not efficient to encode object saliency. To overcome these disadvantages, we describe a novel measurement, namely Corner-Surround Contrast (CSC)*, to segment salient regions from backgrounds. To explore the effects of CSC feature, a kernel-based fusing framework is designed to produce the saliency map automatically and infer the binary segmentation using graph cut algorithm. The experiments demonstrate the promising performance of our method in terms of segmentation accuracy and saliency localization.*

1. Introduction

Saliency detection is a challenging problem in computer vision, and it is also a crucial task in many applications, such as object recognition [2], image coding [3], image editing [4], image segmentation [5], and video tracking [12]. Through lots of efforts, many successful saliency detection methods have been developed, and they are roughly classified into two categories: 1)Top-down (supervised) methods: they often describe the salient information by the visual knowledge constructed from the training process, and then use such knowledge for saliency detection on the test images [7]. 2) Bottom-up (unsupervised) methods: they usually determine the saliency of a pixel based on how it is different from its surrounding neighborhood without any prior of the salient region or object [11]. As one of the earliest methods, Itti *et al.* [8] proposed a center-surround contrast as bottom-up operation in the color, intensity, and orientation of an image. Mutual information is also used to compute the saliency based on discriminant

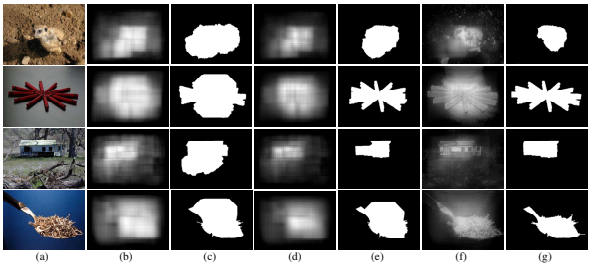

Figure 1. From left to right: (a) original images; (b), (d) and (f) the saliency maps using CSO, CSO+CSC, and our multi-cue fusion framework; (c), (e) and (g) corresponding binary segmentation results.

center-surround hypothesis [6]. Most previous visual attention methods, especially the bottom-up approaches, prefer to use center-surround operation (CSO) to produce better saliency maps. However, this hypothesis often fails when the contrast can not provide enough discrimination between center and surrounding regions. As shown in Figure 1(b) and (c), using such contrast may take the backgrounds as salient areas, which results in inaccurate binary segmentation using graph cut algorithm [10].

To overcome these disadvantages, we present a novel local operation, namely *corner-surround contrast (CSC)*, to measure the saliency for each pixel. As shown in Figure 1(d), after CSO is implemented for one pixel, CSC is then added as supplementary feature to highlight salient objects. To further explore the effect of proposed CSC, we also design a multi-cues integration framework using multiple kernel learning (MKL) [14]. After the saliency maps have been computed based on different features, they are combined linearly with learned weights to produce a final saliency map.

Compared with previous methods, the main contributions of our approach lie in two aspects: firstly, we

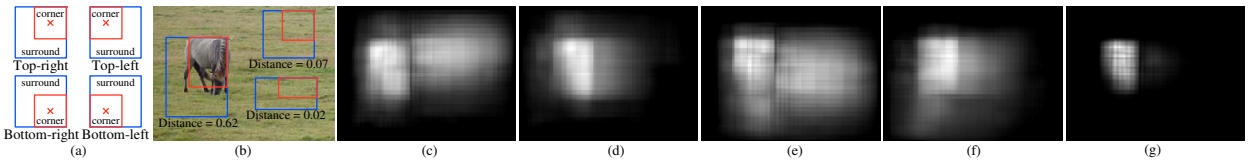

Figure 2. From left to right: (a) four types of CSC operators; (b) top-right CSC distances with different locations and sizes. (c)∼(f) feature maps generated by top-right, top-left, bottom-right, and bottom-left CSCs, respectively. (g) final CSC saliency map. (best viewed in color)

calculate pixel saliency based on CSC feature, which helps us to get more exact location of the salient regions; secondly, we design a cue-fusing method to obtain robust saliency map, which is benefit for generating precise salient segmentation with less noises from the backgrounds.

2. Our Approach

2.1. Corner-surround Contrast (CSC)

As shown in Figure 1(a), the salient object usually has distinct extent, and can be distinguished from its surrounding context. As shown in Figure 2(a), suppose the salient object is enclosed by a red rectangle R. We construct four surrounding blue regions R_s with the same area of R, and define the contrast between each pair of R and R_s as *corner-surround contrast (CSC)*. Based on the relative location between R and R_s, we call these CSCs as *top-left, top-right, bottom-left* and *bottom-right* CSC, respectively.

For one specific CSC (e.g., top-right CSC), to measure how distinct the salient object in the rectangle R is with respect to its surroundings R_s, we can compute the distance between R and R_s based on various visual low-level cues such as intensity, color, and texture/texton. In this paper, we employ the χ^2 distance between histograms of R and R_s in RGB color space: $\chi^2(R, R_s) = \frac{1}{2}\sum_{b=1}^{B}\frac{|R^b - R_s^b|^2}{R^b + R_s^b}$, where $B = 64$ is the number of bins for each histogram. The usage of histograms is robust to describe appearance variance, which is not sensitive to small changes in size, shape, and viewpoint. Another merit lies in that the histogram of a rectangle with any location and size can be computed very quickly using an integral histogram introduced in [13]. Figure 2(b) shows that the salient object (the horse) is most distinct using the χ^2 histogram distance of top-right CSC.

To address the varying aspect ratios of the object, we use five templates with different aspect ratios $\{0.5, 0.75, 1.0, 1.5, 2.0\}$. We find the most distinctive rectangle, $R^*(x)$, centered at pixel x (marked as red "$\times$" shown in Figure 2(a)) by varying the aspect ratio:

$$R^*(x) = \arg\max_{R(x)} \chi^2(R(x), R_s(x)) \quad (1)$$

We also sample the size of rectangle $R(x)$ to handle the scale variance of saliency objects. In practice, this range is set to $[0.1, 0.7] \times min(w, h)$ with range step 0.05, where w, h indicate the image width and height. Then top-right CSC feature $f^{tr}(x)$ is defined as the sum of spatially weighted distances:

$$f^{tr}(x) \propto \sum_{\{x'|x \in R^* x'\}} \omega_{xx'} \chi^2(R^*(x'), R_s^*(x')) \quad (2)$$

where $R^*(x')$ is the rectangle centered at x' and containing the pixel x. $\omega_{xx'} = \exp\{-0.5\sigma_{x'}^{-2}|x - x'|^2\}$ is a Gaussian falloff weight with variance $\sigma_{x'}^2$, which is set to one-third of the size of $R^*(x')$.

Similar with $f^{tr}(x)$, we can computer other three feature maps for each pixel, and denote them as $f^{tl}(x)$, $f^{bl}(x)$ and $f^{br}(x)$, respectively. Figure 2(c)∼(f) show the feature maps generated by corresponding CSC. It is observed that besides highlight salient regions, inducing such CSC operator also gives high scores for backgrounds. As a result, we define the saliency map $sal^{CSC}(x)$ by multiplying four feature maps together:

$$sal^{CSC}(x) = f^{tl}(x) \cdot f^{tr}(x) \cdot f^{bl}(x) \cdot f^{br}(x) \quad (3)$$

The saliency map is normalized to a fixed range [0, 1]. Figure 2(g) illustrates the CSC saliency map, where the salient objects are well located.

2.2. Center-surround Operation (CSO)

We also use the CSO [11] feature to compute saliency for pixels. More specifically, we calculate the histogram for each color component, normalize them, and then concatenate them to get a long histogram. Feeding this feature into CSO operator to obtain the color-based CSO saliency map $sal^{CSO}(x)$. Since CSC and CSO

are both local contrast, they are added together, and we define the local saliency $sal^{LOC}(x)$ as:

$$sal^{LOC}(x) = sal^{CSO}(x) + sal^{CSC}(x) \quad (4)$$

where $sal^{CSO}(x)$ and $sal^{CSC}(x)$ are with equal wights to achieve best performance.

2.3. Color Spatial Variation (CSV)

Unlike CSC and CSO descriptor that investigate local contrast, CSV [11] presents a global feature to measure saliency and has been proved quite efficient.

First, all colors in the image are represented by Gaussian Mixture Models (GMMs) $\{\omega_c, \mu_c, \Sigma_c\}_{c=1}^{C}$, where $\{\omega_c, \mu_c, \Sigma_c\}$ denotes the weight, the mean color, and the covariance matrix of the c^{th} cluster, respectively. $C = 10$ is the number of clusters in our experiments. Each pixel x is assigned to a color cluster with the probability $p(c|I_x) = \frac{\omega_c \mathcal{N}(I_x|\mu_c,\Sigma_c)}{\sum_c \omega_c \mathcal{N}(I_x|\mu_c,\Sigma_c)}$.

To evaluate the saliency of each cluster, we use cluster compactness as a measurement. Intuitively, background clusters tend to have rather larger spread compared to salient clusters. Therefore, the larger the spatial variance is, the less compactness this cluster has: $V(c) = \frac{\sum_x p(c|I_x)|x-\mu_c|^2}{\sum_x p(c|I_x)}$, where μ_c is the spatial mean of the c^{th} cluster, and $V(c)$ is normalized to [0, 1]. The saliency map $sal^{CSV}(x)$ of pixel x is defined as:

$$sal^{CSV}(x) \propto \sum_c p(c|I_x) \cdot (1 - V(c)) \quad (5)$$

2.4. Multi-scale Contrast (MSC)

Without prior knowledge about the size of the salient object, contrast is usually computed at multiple scales. Following [11], we simply define the MSC saliency map for pixel x using Gaussian image pyramid:

$$sal^{MSC}(x) = \sum_{l=1}^{L} \sum_{x' \in N_x} |I_x^l - I_{x'}^l|^2 \quad (6)$$

where I^l is the l^{th} level image in the pyramid and the number of pyramid levels L is 6. N_x is a 9×9 window. This saliency map is normalized to a fixed range [0, 1].

2.5. Saliency Map Combination

Intuitively, different features may have different proportion of contribution for final saliency map. In addition, although we have established and normalized the saliency maps using CSC, CSO, CSV and MSC features, these maps are always statistical fluctuations within different feature spaces. One solution is using MKL scheme [14], which defines the final saliency map $sal^{fn}(x)$ by linearly integrating all saliency maps with different weights:

$$sal^{fn}(x) = \sum_{n=1}^{N} d_n K_n(sal(x), sal(x')) \quad (7)$$

where $N = 3$ is the number of saliency maps, $sal(x)$ is the saliency map defined in Equation 4, Equation 5, and Equation 6, respectively. $K_n(sal(x), sal(x')) = \exp\{|sal(x) - sal(x')|\}$ is a radial basis kernel, and d_n indicates the feature weight.

In [14], the authors demonstrate the formulation of Equation 7 is actually the dual of a support vector machine (SVM) problem, which can be solved efficiently through a primal formulation involving a weighted ℓ_2-norm regularization. Another advantage of using [14] is the weights are always normalized: $\sum_n d_n = 1, d_n > 0$. Employing MKL method produces the robust results with all features, as shown in the Figure 1(f) and Figure 4. The best linear weights we learnt are: $\{0.58, 0.25, 0.17\}$.

3. Experiments

In this section, we evaluate the performance of our method with existing methods on Microsoft Research Asian (MSRA) [1] dataset including 1000 images, which provides accurate object-contour-based ground truth. For utilizing the MKL to train the weights in Equation 7, all images are randomly split into roughly 40% for training, and 60% for testing. To show the advantages of our approach, we employ four state-of-the-art models for comparison, namely, IT [8], FT [1], CA [7] and GB [9], and implement all methods using a Dual Core 2.6 GHz machine with 4GB memory.

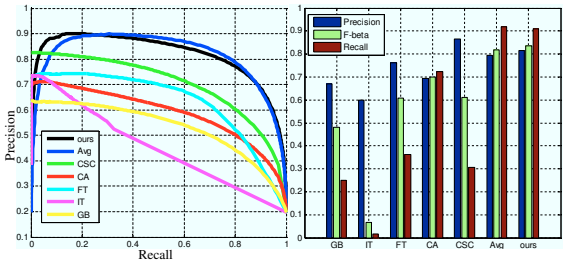

Figure 3. P-R curves and P-R bars. (best viewed in color)

Quantitative results. To compare how well various saliency detection methods highlight salient regions, we

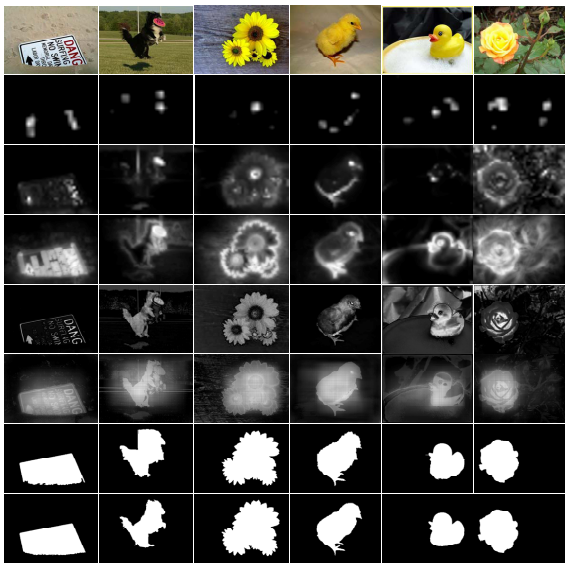

Figure 4. From top to bottom: original images, saliency maps produced by IT [8], GB [9], CA [7], FT [1] and ours. Final two rows are our binary segmentation and ground truth.

vary a threshold T_f from 0 to 255 on obtained saliency maps. Left panel of Figure 3 shows the precision vs. recall (P-R) curves. We also address the contribution of individual CSC feature, and present the benefits of using MKL weighting schema, along with objective comparison with equal weighting schema (denoted as Avg) and other saliency extraction methods.

Another criterion to evaluate the overall performance is *F-measure*, which is computed with non-negative β:

$$F_\beta = \frac{(1+\beta) \times Precision \times Recall}{\beta \times Precision + Recall} \qquad (8)$$

where $\beta = 0.3$ following [4]. Right panel of Figure 3 gives the *F-measure* evaluation results. As can be seen, our method shows high precision, recall, and F_β values over the 1000-image database.

Qualitative results. Once given the final saliency map, we infer the binary segmentation using graph cut algorithm [10]. Figure 4 exhibits the saliency maps of some examples obtained by the various methods for visual comparison. Our method can highlight salient regions, and generate more accurate binary segmentation results.

4. Conclusion

In this paper, a novel contrast, namely CSC, is introduced to capture local dissimilarity of salient objects. In addition to this bottom-up setting, the proposed CSC measurement can be also generalized to incorporate the top-down priors obtained from MKL algorithm. Extensive experiments well validate the effect of CSC feature on the MSRA dataset. Compared with existing methods, our approach gains significantly in terms of segmentation accuracy and saliency localization. Another advantage of our CSC operator lies in its flexibility, which is adaptive to new features. In the future, we should extend to explore texture features, which may make our method more robust in the cases where there is no dominant color in the image.

Acknowledgment

This work was supported by the National Natural Science Foundation of China, grant No. 61173120.

References

[1] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. In *CVPR*, pages 1597–1604, 2009.

[2] B. Alexe, T. Deselaers, and V. Ferrari. What is an object? In *CVPR*, pages 73–80, 2010.

[3] A. Bradley and F. Stentiford. Visual attention for region of interest coding in jpeg 2000. *JVCIR*, 14(3):232–250, 2003.

[4] M. Cheng, G. Zhang, N. Mitra, X. Huang, and S. Hu. Global contrast based salient region detection. In *CVPR*, pages 409–416, 2011.

[5] Y. Fu, J. Cheng, Z. Li, and H. Lu. Saliency cuts: An automatic approach to object segmentation. In *ICPR*, pages 1–4, 2008.

[6] D. Gao and N. Vasconcelos. Bottom-up saliency is a discriminant process. In *ICCV*, pages 1–6, 2007.

[7] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. In *CVPR*, pages 2376–2383, 2010.

[8] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *TPAMI*, 20(11):1254–1259, 1998.

[9] H. J., K. C., and P. P. Graph-based visual saliency. In *NIPS*, pages 545–552, 2006.

[10] V. Kolmogorov and R. Zabin. What energy functions can be minimized via graph cuts? *TPAMI*, 26(2):147–159, 2004.

[11] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Shum. Learning to detect a salient object. *TPAMI*, 33(2):353–367, 2011.

[12] V. Mahadevan and N. Vasconcelos. Spatiotemporal saliency in dynamic scenes. *TPAMI*, 32(1):171–177, 2010.

[13] F. Porikli. Integral histogram: A fast way to extract histograms in cartesian spaces. In *CVPR*, pages 829–836, 2005.

[14] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet. Simplemkl. *JMLR*, 9:2491–2521, 2008.