Salient Region Detection with Convex Hull Overlap

Yongqing Liang  Cheng Jin  Yuejie Zhang

root@lyq.me  jc@fudan.edu.cn  yjzhang@fudan.edu.cn

ABSTRACT
In this paper, we establish a novel bottom-up cue named Convex Hull Overlap (CHO), and then propose an effective approach to detect salient regions using the combination of the CHO cue and global contrast cue. Our scheme significantly differs from other earlier work in: 1) The hierarchical segmentation model based on Normalized Graph-Cut fits the splitting and merging processes in human visual perception; 2) Previous work only focuses on color and texture cues, while our CHO cue makes up the obvious gap between the spatial region covering and the region saliency. CHO is a kind of improved and enhanced Gestalt cue, while other popular figure-ground cues such as convexity and surroundedness can be regarded as the special cases of CHO. Our experiments on a large number of public data have obtained very positive results.

CCS Concepts
• Computing methodologies—Artificial intelligence—Computer vision—Computer vision problems—Interest point and salient region detections.

Keywords
Salient region detection; Convex Hull Overlap (CHO); hierarchical segmentation; saliency map; Gestalt convexity.

1. INTRODUCTION
Usually, humans can detect visually distinctive noticeable foregrounds in the scene (that is, salient objects or regions) effortlessly and rapidly. This capability has long been studied by cognitive scientists and attracted a lot of research interests for salient region detection in the field of computer vision. Previous studies on salient region detection focus on two main strategies, in which the top-down pattern is slow, task-driven and determined by cognitive phenomena like knowledge, expectations, reward and current goals, and the bottom-up pattern indicates that the selection of salient regions depends on low-level features (e.g., color, texture, geometry, overlap, etc.) of images. At present, most related research work has done well on the bottom-up-based pattern [1][4][7][12][18]. Such a pattern generally utilizes the regular cues to exhibit the visual properties of image contents, which is effective for salient region detection but involves the potential need of the in-depth valuable cue discovery for more accurate detection. Thus, the salient region detection with more comprehensive cue information is becoming more popular to alleviate the above problems, which has been proved to be successful in more recent related work.

Although salient region detection has been extensively studied since recent years, it still remains the necessity of optimal solutions and three interrelated issues should be addressed simultaneously: 1) an efficient model to accurately represent human visual perception for the multi-scale analysis of objects and scenes; 2) an appropriate feature representation to bridge the gap between the spatial covering and the saliency; and 3) the in-depth figure-ground analysis to improve the result of salient region detection. Meanwhile, in cognition science the research on salient region detection especially the figure-ground analysis has been beneath the attention of Gestalt psychologists like Wertheimer et al., since 1920s. Historically, Gestalt psychology has emphasized that the Gestalt laws are innate and intrinsic to the brain rather than learned from past experience. Convexity, one of the Gestalt laws, has been proved as a significant bottom-up cue to separate figural and background regions in perceptual organization. The convexity rule suggests that the region on the convex side of a curved boundary tends to be a figure.

Based on the above observations, we establish a novel bottom-up cue named Convex Hull Overlap (CHO), and then propose an effective approach to detect salient regions using the combination of the CHO cue and global contrast cue, as shown in Figure 1. The hierarchical segmentation model is firstly constructed by using the recursive Normalized Graph-Cut algorithm [15] to express the visual perception for objects and scenes more precisely. The convex hull is defined as the smallest convex set that contains all the pixels of a region in the Euclidean plane, and then CHO can be computed as the figural confidence of regions. Finally, the new feature description with CHO cues can be effectively utilized to separate figural and background regions. Our scheme significantly differs from other earlier work in: 1) The hierarchical segmentation model based on Normalized Graph-Cut fits the splitting and merging processes in human visual perception; 2) Previous work only analyzes with color and texture cues, while our CHO cue makes up the obvious gap between the spatial region covering and the region saliency. CHO is a kind of improved and enhanced Gestalt cue, while other popular figure-ground cues such as convexity and surroundedness can be regarded as the special cases of CHO. We take advantage of the CHO cue and the traditional global contrast cue to generate a novel mixed cue. Our experiments on a large number of public data have obtained very positive results.

2. RELATED WORK
Salient region detection belongs to the active research field of visual attention modeling. A major distinction among visual attention models is whether they rely on bottom-up influences, top-down influences, or a combination of both. Because bottom-up cues are fast, exogenous, automatic and involuntary [3], the increasing research interests focus on such cues.
In recent years, intensity and color contrast features/cues are simple and widely used in salient region detection. Zhai et al. used luminance information to seek interesting regions in images [18]. Achanta et al. defined the pixel saliency using the pixels’ color difference from the average image color [1]. Cheng et al. computed the saliency map by evaluating global color contrast differences among regions [4]. Rahtu et al. adopted illumination, color and motion information with a Conditional Random Field (CRF) model [13]. Spectral information was utilized in Hou et al. [7], in which they analyzed the log-spectrum of an input and proposed a fast method to construct the corresponding saliency map in the spatial domain. Lu et al. exploited convexity and surroundedness cues, both of figure-ground cues, to detect salient objects, but their model was sensitive to the super-pixels boundaries [12]. Jiang et al. integrated both the bottom-up salient stimuli and the object-level shape prior [8]. Zhou et al. integrated compactness and local contrast cues to detect salient regions [19], but it may fail for images that do not have much color variation. Kim et al. composed an accurate saliency map by finding the optimal linear combination of color coefficients in the high-dimensional color space [9]. Scharfenberger et al. proposed a structure-guided statistical textural distinctiveness approach to salient region detection [14], Zou et al. presented a novel unsupervised algorithm using Markov random field [20], but their method is very computationally expensive.

From another perspective of neuroscience and psychology, there exist an amount of research work on finding different ways to explore more beneficial cues to illustrate the obvious appearances for salient regions through the chosen effective cues. Wagemans et al. concluded that the connectivity and rules of visual cortex allowed the illusory contours to be formed and the figure-ground segmentation to be performed by the autonomous processes [16]. Kimchi et al. claimed that the figure-ground segregation could occur before the focal attention [10]. It’s worth noting that the convexity cue of Gestalt laws, as widely believed, is a strong cue to assign the figure-ground relation. Bertamini et al. provided the evidence that the visual system could extract the information about convexities and concavities along with contours in an image [2]. They concluded that the convexity affected the figure-ground relation. Fowlkes et al. found that the convexity indeed had the ability to discriminate which part was the foreground on natural images by ecological statistics [6].

Unfortunately, all these existing methods have not yet provided good solutions for the following two crucial issues: 1) Improving the hierarchical segmentations -- Most existing methods build the hierarchical segmentation model by varying the number of super-pixels using the segmentation algorithm in Felzenswalb et al. [5]. After segmentation, there is only one connected component for any region. However, in reality an occluder may separate two regions from the same object. Thus in some cases, it’s necessary to merge such two regions that belong to the same object into one whole region. 2) Developing affluent figure-ground cues for detection -- In most cases, the color contrast and spatial information are sufficient to evaluate salient regions. Nevertheless, in the complex background, the background regions may be labeled as salient regions because of the sharp color contrast. The figure-ground cue, one of the bottom-up cues, can label regions either foreground or background based on their shapes, while the robust cue is underutilized for a long time. Thus it’s very imperative to construct a novel figure-ground cue to bridge the gap between the figure-ground cue and region saliency, especially under the overall consideration of neuroscience and psychology.

3. HIERARCHICAL SEGMENTATION

This is the preprocessing process in our algorithm. Our goal is to build n hierarchical segmentations from the input image. In any region, the color values of pixels are similar and the space positions of pixels are close but do not need to be adjacent to each other. We first segment the input image into regions with the over segmentation, and then build the hierarchical model using the Normalized Graph-Cut algorithm [15].

The input image is segmented into the regions of \( \{r_1, r_2, \cdots, r_m\} \) by using the over segmentation based on Felzenswalb et al.’s algorithm [5], as shown in Figure 2(b). \( L' a' b' \) is a color-opponent space where the dimension \( L' \) is for lightness, the dimension \( a' \) is the position between red/magenta and green, and the dimension \( b' \) is the position between yellow and blue. A weighted graph \( G = (V, E) \) is constructed by taking each region as a node and connecting each pair of regions by an edge. The weight on the edge should reflect the likelihood that two regions belong to one object. With the \( L' a' b' \) value of the pixels and their spatial locations, the weight for the graph edge connecting two nodes of \( r_i \) and \( r_j \) can be defined as:

\[
w(r_i, r_j) = \left( |r_i| \cdot |r_j| \right) \cdot e^{-\frac{D_L(r_i, r_j)}{\sigma_L^2}} \cdot e^{-\frac{D_a(r_i, r_j)}{\sigma_a^2}}
\]  

(1)

where \( D_L(r_i, r_j) \) is the \( L' a' b' \) space distance between two regions of \( r_i \) and \( r_j \); \( \sigma_L^2 \) aims to control the strength of the color distance weighting; \( D_a(r_i, r_j) \) is the spatial distance between two regions of \( r_i \) and \( r_j \); \( \sigma_a^2 \) aims to control the strength of the spatial distance weighting; and \( |r_i| \) is the amount of pixels of Region \( r_i \).

The Normalized Graph-Cut algorithm is recursively utilized to construct \( n \) hierarchical segmentations (in our experiments the number of layers is set as 5), as shown in Figure 2(c). We define \( L = \{l_1, l_2, \cdots, l_n\} \) as the set of hierarchical segmentations, in which \( l_1 \) is the coarsest segmentation and \( l_n \) is the finest one.

4. SALIENCY MAP GENERATION

In our algorithm, we integrate both the CHO cue and global contrast cue to generate more accurate saliency map \( S \):

\[
S(p) = S_{CHO}(p) \cdot S_{GC}(p)
\]

(2)

where \( p \) is a pixel point of image. Firstly, we compute the Convex Hull Overlap map \( S_{CHO} \) and the global contrast map \( S_{GC} \) separately. Secondly, we merge two maps to produce the pixel-level saliency map.
4.1 Convex Hull Overlap Map

In our approach, the region \( R \) consists of a set of pixel points \( P = \{p_1, p_2, \ldots, p_m\} \). A convex polygon, whose interior angles are less than 180 degrees, is a finite region that contains all the pixel points. The convex hull \( C \) of \( R \) is the smallest convex polygon that contains \( P \). Because \( R \) is often an irregular shape, \( C \) is the corresponding circumscribed convex polygon. The gap between \( R \) and \( C \) is filled with the pixels belong to the other regions, and we define the gap as Convex Hull Overlap (CHO). We use Graham’s Scan algorithm to find the convex hull of a finite set of points in the plane with the time complexity \( O(m\log m) \).

We define the set of regions in Layer \( L_i \) as \( \{R^i_1, R^i_2, \ldots, R^i_n\} \), and compute the convex hull \( C^i_j \) for each region \( R^i_j \) in \( L_i \). Based on the definition above, \( C^i_j \) contains all the pixels of \( R^i_j \). Besides that, \( C^i_j \) may also contain the pixels of the other regions. Obviously, on the one hand, if a region is the background, its convex hull may contain several pixels belong to the other regions. On the other hand, if a region is the foreground, it may occupy the pixels of the other convex hulls. We define the convex hull overlap of \( R^i_j \) as:

\[
\text{CHO}(R^i_j) = \frac{\text{all other convex hulls } C | C \cap R^i_j |}{| R^i_j |} \tag{3}
\]

where \( | R^i_j \cap C | \) is the amount of pixels that are marked as \( R^i_j \) and contained by \( C \); and \( | R^i_j | \) is the amount of \( R^i_j \).

We compute the CHO saliency of each pixel as follows:

\[
S_{\text{CHO}}(p) = \frac{1}{n} \sum_{i=1}^{n} \text{CHO}(R^i_j), p \in R^i_j \tag{4}
\]

where \( n \) is the amount of layers. Figure 3 shows the details for the CHO cue evaluation, and Figure 3(c) shows the final CHO map, in which the bright regions indicate their saliency.

4.2 Global Contrast Map

As is known to all, humans are highly sensitive to the color contrast in the visual signals. We define the color contrast and then further incorporate the spatial information by introducing a spatial weighting term to increase the effects of closer regions and decrease the effects of farther regions. The global contrast of any region \( R^i_j \) in Layer \( L_i \) is defined as follows:

\[
S_{\text{GC}}(R^i_j) = w_s(R^i_j) \sum_{\text{all regions } R^i_k \text{ in } L_i} \frac{D_c(R^i_j, R^i_k)}{\sigma^2} \tag{5}
\]

where \( w_s(R^i_j) \) is given a low value if \( R^i_j \) is a border region far from the center and given a high value if the region is close to the center of the image; \( D_c(R^i_j, R^i_k) \) is the spatial distance between two regions of \( R^i_j \) and \( R^i_k \); \( \sigma^2 \) is the spatial distance between \( R^i_j \) and \( R^i_k \); and \( \sigma^2 \) aims to control the strength of the spatial distance weighting.

We compute the global contrast saliency of each pixel as follows:

\[
S_{\text{GC}}(p) = \frac{1}{n} \sum_{i=1}^{n} S_{\text{GC}}(R^i_j) \cdot p \tag{6}
\]
because the color information does not work well in the complex background, while the geometrical and covering information can be still effective to distinguish the figural and background regions. It also indicates that the in-depth figure-ground analysis by using our approach with the CHO cue is necessary and efficient to improve the salient region judgement.

Table 1. The quantitative comparison analysis between the related state-of-art methods and ours with the best F-measure rate.

| Dataset     | Metric | Approach |
|-------------|--------|----------|
|             |        | Ours | RC | HC | LC | SR | FT |
| PASCAL-S     | Precision | 0.84 | 0.77 | 0.72 | 0.57 | 0.45 | 0.72 |
|              | Recall  | 0.77 | 0.64 | 0.66 | 0.44 | 0.45 | 0.55 |
|              | F-measure | 0.83 | 0.74 | 0.70 | 0.53 | 0.45 | 0.67 |
| ECSSD        | Precision | 0.78 | 0.72 | 0.48 | 0.42 | 0.40 | 0.45 |
|              | Recall  | 0.63 | 0.62 | 0.46 | 0.38 | 0.45 | 0.45 |
|              | F-measure | 0.74 | 0.70 | 0.47 | 0.41 | 0.41 | 0.45 |
| MSRA10K      | Precision | 0.89 | 0.83 | 0.72 | 0.62 | 0.51 | 0.70 |
|              | Recall  | 0.73 | 0.62 | 0.63 | 0.50 | 0.49 | 0.55 |
|              | F-measure | 0.85 | 0.77 | 0.70 | 0.59 | 0.50 | 0.65 |

Although our approach can detect the most salient objects accurately, it still has some limitations in some particular cases. For example, if the foreground and background have exactly the same color, the segmentation fails to separate them into two regions and our result is undesirable. This may be the stubbornest problem for salient region detection.

6. CONCLUSION

A new framework is introduced in this paper to support more precise salient region detection for large-scale images. The Convex Hull Overlap cue is especially introduced to achieve an optimal representation for images. The mixed cue with the integration of the CHO cue and global contrast cue is particularly established for characterizing the meaningful visual information for prominent regions in images more precisely. Different from machine learning, our scheme using hand-crafted feature has advantages in the results and computational consuming. Our future work will focus on setting up a specialized learning mechanism to implement the deep-level analysis and mining for in-depth cues.
7. REFERENCES

[1] Achanta, R., Hemami, S., Estrada, F., & Susstrunk, S. (2009, June). Frequency-tuned salient region detection. In Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on (pp. 1597-1604). IEEE.

[2] Bertamini, M., & Wagemans, J. (2013). Processing convexity and concavity along a 2-D contour: figure–ground, structural shape, and attention. Psychonomic bulletin & review, 20(2), 191-207.

[3] Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 35(1), 185-207.

[4] Cheng M, Mitra N J, Huang X, et al. Global contrast based salient region detection[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2015, 37(3): 569-582.

[5] Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. International Journal of Computer Vision, 59(2), 167-181.

[6] Fowlkes, C. C., Martin, D. R., & Malik, J. (2007). Local figure–ground cues are valid for natural images. Journal of Vision, 7(8), 2-2.

[7] Hou, X., & Zhang, L. (2007, June). Saliency detection: A spectral residual approach. In Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on (pp. 1-8). IEEE.

[8] Jiang, H., Wang, J., Yuan, Z., Liu, T., Zheng, N., & Li, S. (2011, September). Automatic salient object segmentation based on context and shape prior. In BMVC (Vol. 6, No. 7, p. 9).

[9] Kim, J., Han, D., Tai, Y. W., & Kim, J. (2016). Salient Region Detection via High-Dimensional Color Transform and Local Spatial Support. Image Processing, IEEE Transactions on, 25(1), 9-23.

[10] Kimchi, R., & Peterson, M. A. (2008). Figure-ground segmentation can occur without attention. Psychological Science, 19(7), 660-668.

[11] Li, Y., Hou, X., Koch, C., Rehg, J., & Yuille, A. (2014). The secrets of salient object segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 280-287).

[12] Lu, Y., Zhang, W., Lu, H., & Xue, X. (2011, November). Salient object detection using concavity context. In Computer Vision (ICCV), 2011 IEEE International Conference on (pp. 233-240). IEEE.

[13] Rahtu, E., Kannala, J., Salo, M., & Heikkilä, J. (2010). Segmenting salient objects from images and videos. In Computer Vision–ECCV 2010 (pp. 366-379). Springer Berlin Heidelberg.

[14] Scharfenberger, C., Wong, A., & Clausi, D. A. (2015). Structure-guided statistical textural distinctiveness for salient region detection in natural images. Image Processing, IEEE Transactions on, 24(1), 457-470.

[15] Shi, J., & Malik, J. (1997, June). Normalized cuts and image segmentation. In Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on (pp. 731-737). IEEE.

[16] Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. Psychological bulletin, 138(6), 1172.

[17] Yan, Q., Xu, L., Shi, J., & Jia, J. (2013). Hierarchical saliency detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1155-1162).

[18] Zhai, Y., & Shah, M. (2006, October). Visual attention detection in video sequences using spatiotemporal cues. In Proceedings of the 14th annual ACM international conference on Multimedia (pp. 815-824). ACM.

[19] Zhou, L., Yang, Z., Yuan, Q., Zhou, Z., & Hu, D. (2015). Salient Region Detection via Integrating Diffusion-Based Compactness and Local Contrast. Image Processing, IEEE Transactions on, 24(11), 3308-3320.

[20] Zou, W., Liu, Z., Kpalma, K., Roncin, J., Zhao, Y., & Komodakis, N. (2015). Unsupervised joint salient region detection and object segmentation. Image Processing, IEEE Transactions on, 24(11), 3858-3873.