Improvement of Saliency Optimization from Robust Background Detection

Jiayi Fang
University of Michigan, Ann Arbor
fangjy@umich.edu

Wentao Zhang
University of Michigan, Ann Arbor
zwtao@umich.edu

Jiangchen Zhu
University of Michigan, Ann Arbor
zjcsjtu@umich.edu

Haoxiong Su
University of Michigan, Ann Arbor
sowone@umich.edu

Abstract

The method of the salient region detection is critical in the computer vision field. One recent progress made in improving the saliency detection is to use the so called boundary prior or background information. With the use of the boundary prior, traditional methods that utilize other cues are improved effectively. However, in some papers, the use of the boundary prior is too vague or simple to efficiently make the advantage of it. Our work is divided into two parts. First,we replicate the saliency optimization method via the robust background measure, called boundary connectivity. Second, we integrate the boundary prior into the low-level cues of the image to generate the saliency maps. The detection of the salient regions is reframed as the principled optimization problem that is solved by existing methods. Lastly, we try to use the edge detection method to enhance the saliency maps. In the result of the previous work, we observe that if the objects and the background regions have the similar color/contrast, the method cannot distinguish them well. The edge detection is highly useful to solve the problem.

1. Introduction

Salient region detection is essential to understand and analyze the images. The goal is to detect salient regions in the image given the saliency map. The saliency map is the scores of the regions in the image that represent how salient they are. In recent years, salient region detection is more and more important since it is used in many applications, such as image segmentation, object recognition, image cropping, etc. The most widely used assumption in the salient region detection methods is the contrast prior. The assumption states that *appearance contrasts between objects and their surrounding regions are high*. The contrast prior is used in many existing methods like [2] [3].

Another prior widely used in the detection of the salient region is the boundary prior. The assumption states that *image boundary regions are mostly backgrounds*, to enhance saliency computation. The prior is proved to be effective in some methods [6]. However, there are still some obvious drawbacks of them. In some methods, they simply treat all the boundaries of the image as the background. The assumption made here is problematic when the object in the image stays around the boundary. Besides, for the existing methods the integration of the boundary prior into the low-level curs in unclear. In general, the integration is heuristics.

In our project, we are inspired of using the background measure, called boundary connectivity, in the paper [8]. In the above, we talked about the existing problem of simply treating the boundary of the image as the background. Instead of doing so, the boundary connectivity accounts for the pixels that are easily connected to the image boundary as the background. The background measure assumes that *an image patch is background only when the region it belongs to is heavily connected to the image boundary*. As stated by the authors, this measure is more robust as it characterizes the spatial layout of image regions with respect to image boundaries. One advantage is that using the boundary connectivity could easily handle the images with only the backgrounds but the objects. The boundary connectivity could address the problem above significantly by setting an appropriate threshold value for the measuring how easily the pixels are connected to the boundary.

To detect the salient regions of the image, some methods aggregate multiple low-level cues of the pixels to determine the saliency. Generally, the low-level cues are combined by the weighted sum or the multiplication. In our project, we also replicate the framework that transforms the estimation problem as the principled optimization problem. The cost function is defined aiming for the goal of generating the true saliency maps. The objects are constrained to take the high saliency scores. In the reverse, the backgrounds are constrained to have the low saliency scores. With the distinguished scores between the objects and the backgrounds in the saliency maps, we could easily detect the salient regions. To address the problem of the transition from the backgrounds to the objects, we have integrated the smoothness constraints into the optimization problem as well. All the constraints used in the optimization problem are linear. With the use of the low-level cues, the method could detect the salient regions of the image efficiently and straightforward.

In the result of the previous work, we notice that the method does not perform well on the images where the objects and the backgrounds share the similar colors/contrast. Therefore, we are inspired to use the edge detection on the initial saliency maps to extract the highly distinctive regions as the objects. The result shows that after the edge detection, the saliency maps are more accurate without mistakenly taking the backgrounds as the object frequently.

2. Related work

Some methods of the salient region detection utilize the color values based on the observation that distinctive colors compared to the background capture more attention from the human perception. Under this assumption, they utilize the color features of the regions to distinguish between the objects and the backgrounds. To detect the salient regions, the paper [5] generates the feature vectors and apply the high-dimensional transformation to them. The method is effective but still has some drawbacks. For example, if the object has the similar color with the one of the background, the method fails to distinguish between them in the high-dimensional space.

The edge detection is useful for finding the boundaries of objects within images. We also take the use of the edge detection to improve our saliency maps by extracting the salient segmentation as the objects. The methods of the edge detection are various [7]. Generally speaking, the edge detection focus on the discontinuity of the images. This method is useful to us since there are discontinuous transition between the objects and the backgrounds in the initial saliency maps.

In this project, we also refer to the paper [8]. The paper introduces the concept of the background measure, called boundary connectivity. The measure is to address the problem occurred while simply using the background prior to detect the salient regions in most of the projects. Besides, we replicate their framework of transforming the estimation problem into the optimization problem with multiple constraints. The method is efficient and straightforward.

3. Boundary Connectivity

Based on the assumption that *an image patch is background only when the region it belongs to is heavily connected to the image boundary*, we derive the background measure first, called boundary connectivity.

3.1. Conceptual Definition

The authors of [8] observes that the object and background regions in natural images are quite different in their spatial layout, i.e., object regions are much less connected to image boundaries than background ones. From the original paper, there is an example of the synthetic image to explain that as shown in figure 1.

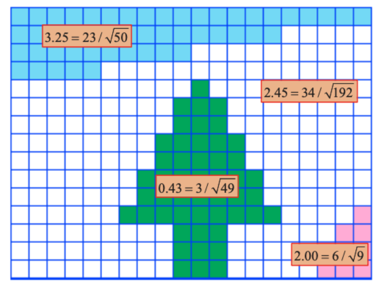


Figure 1. (Better viewed in color) An illustrative example of boundary connectivity. The synthetic image consists of four regions with their boundary connectivity values (Eq.(1)) overlaid. The boundary connectivity is large for background regions and small for object regions.

The synthetic natural image consists of two objects, including the green tree at the center and the pink object that touches the boundary. The blue region represent the sky that should be considered as part of the background. And the other region is also considered as the background.

The proposed boundary connectivity is measuring how heavily a region R is connected to the image boundaries. It is calculated as follows:

$$\text{BndCon}(R) = \frac{|\{p|p \in R, p \in \text{Bnd}\}|}{\sqrt{|p|p \in R|}} \quad (1)$$

where Bnd is the set of image boundary patches and p is an image patch. The measure is intuitive since its value represents *the ratio of a region's perimeter on the boundary to the region's overall perimeter, or square root of its area*.

In the example of figure 1, we observe that the BndCon of the green tree is much lower than the others. Since the pick object is touching the boundary, its BndCon is a little bit higher but still lower than the ones of the backgrounds. The example clearly shows that the boundary connectivity could be used to capture the differences between the object regions and the background regions effectively.

3.2. Effective Computation

Eq.(1) is intuitive but hard to compute because image segmentation itself is a challenging and unsolved problem.

However, in this case, image segmentation is used to calculate boundary connectivity and it does not need to be very accurate. Thus we are able to provide a computationally effective method to approximately approach image segmentation.

We first abstract each image as a set of nearly regular superpixels using the SLIC method [1]. We then construct an undirected weighted graph by connecting all adjacent superpixels (p, q) and assigning their weight $d_{app}(p, q)$ as the Euclidean distance between their average colors in the CIE-Lab color space. The geodesic distance between any two superpixels $d_{geo}(p, q)$ is defined as the accumulated edge weights along their shortest path on the graph

$$d_{geo}(p, q) = \min_{p_1=p,p_2,\dots,p_n=q} \sum_{i=1}^{n-1} d_{app}(p_i, p_{i+1})$$

The shortest paths between all superpixel pairs are efficiently calculated using Johnson's algorithm [4]. And we define $d_{geo}(p, p) = 0$. Then we define the "spanning area" of each superpixel p as

$$Area(p) = \sum_{i=1}^{N} exp(-\frac{d^2_{geo}(p, p_i)}{2\sigma^2_{clr}}) = \sum_{i=1}^{N} S(p, p_i), \quad (2)$$

where N is the number of superpixels.

Eq.(2) computes a soft area of the region that p belongs to. To see that, we note the operand $S(p, p_i)$ in the summation is in (0, 1] and characterizes how much superpixel p_i contributes to p's area. As analyzed in [8], when p_i and p are in a flat region, $d_{geo}(p, p_i) = 0$ and $S(p, p_i) = 1$, ensuring that p_i adds a unit area to the area of p. When p_i and p are in different regions, there exists at least one strong edge ($d_{app}(*, *) \gg 3\sigma_{clr}$) on their shortest path and $S(p, p_i) \approx 0$, ensuring that p_i does not contribute to p's area.

We define the *length along the boundary* as

$$Len_{bnd}(p) = \sum_{i=1}^{N} S(p, p_i) \cdot \delta(p_i \in Bnd),$$

where $\delta(\cdot)$ is 1 for superpixels on the image boundary and 0 otherwise.

Finally, we compute the boundary connectivity as

$$BndCon(p) = \frac{Len_{bnd}(p)}{\sqrt{Area(p)}}$$

Since a physically connected background region can be separated due to occlusion of foreground objects, we also add edges between any two boundary superpixels. It enlarges the boundary connectivity values of background regions and has little effect on the object regions.

3.3. Background Weighted Contrast

[8] introduces the extended equation to calculate a superpixel's contrast, called *background weighted contrast*, is defined as

$$wCtr(p) = \sum_{i=1}^{N} d_{app}(p, p_i) w_{spa}(p, p_i) w_i^{bg}, \quad (3)$$

where $w_{spa}(p, p_i) = exp(-\frac{d^2_{spa}(p,p_i)}{2\sigma^2_{spa}})$ and $w_i^{bg} = 1 - exp(-\frac{BndCon^2(p_i)}{2\sigma^2_{bndCon}})$. $d_{spa}(p, p_i)$ is the distance between the centers of superpixel p and p_i. The new weighting term w_i^{bg} is background probability. It is close to 1 when boundary connectivity is large, and close to 0 when it is small.

As said in [8], according to Eq.(3), the object regions receive high w_i^{bg} from the background regions and their contrast is enhanced. On the contrary, the background regions receive small w_i^{bg} from the object regions and the contrast is attenuated. This asymmetrical behavior effectively enlarges the contrast difference between the object and background regions.

4. Saliency Optimization

As stated above, some of the previous works use the heuristic way to combine multiple low-level cues, which is not hard for generalization. To address the problem, we transform the estimation problem into the principled optimization problem. The objective cost function is designed to assign the object region value 1 and the background region value 0, respectively.

Let the saliency values of N superpixels be $\{s_i\}_{i=1}^N$. The cost function is defined as follows:

$$\underbrace{\sum_{i=1}^{N} w_i^{bg} s_i^2}_{\text{background}} + \underbrace{\sum_{i=1}^{N} w_i^{fg} (s_i - 1)^2}_{\text{foreground}} + \underbrace{\sum_{i,j} w_{ij} (s_i - s_j)^2}_{\text{smoothness}}$$

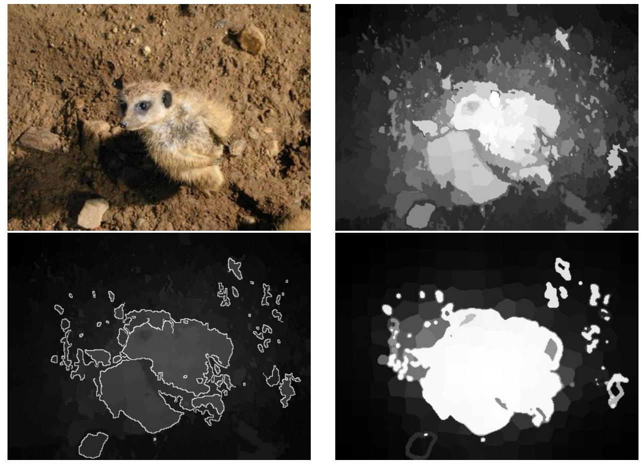
where the three summations indicate the three constraints respectively. The background term encourages a superpixel p_i with large background probability w_{bg} to take a small value s_i (close to 0). In reverse, the foreground term encourages a superpixel p_i with large foreground probability w_{bg} to take a large value s_i (close to 1). The last term, smoothness, encourages continuous saliency values. In this term, the probability is calculated as follows:

$$w_{ij} = exp(\frac{d^2_{\text{app}}(p_i, p_j)}{2\sigma^2_{\text{clr}}}) + \mu$$

The probability is large in the flat regions and small at region boundaries. The reason for it is that the distances between the superpixels are small in the flat regions and similarly at region boundaries.

The optimization problem is solved by the least square method. The solution to the optimization problem is the optimal saliency maps. Generally speaking, the robust background detection performs well on the normal images with high contrast between the objects and the backgrounds.

While using the robust background detection, we found that it performs really bad if the object is in the same color with the background, or the background is messy. This is due to the nature of the salient region detection. Our group finds a solution to remove the noise of the background so that the salient object is shown in a much clearer way.



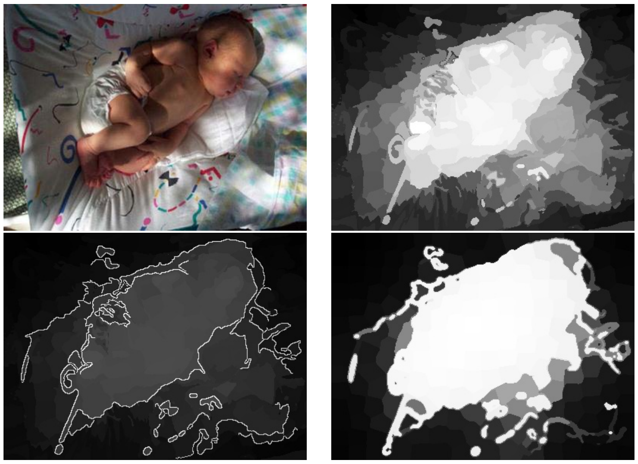
5. Enhancement via Edge Detection

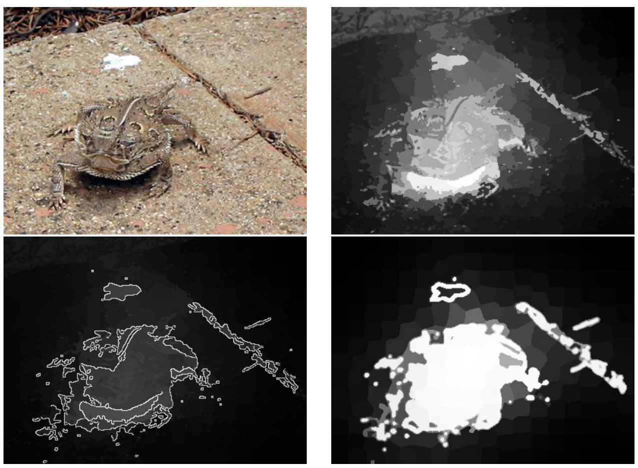
Considering the fact that robust background detection uses smooth boundary as a factor when calculating the saliency cost function, we decided to enhance the edge from the original pictures so that the saliency object will receive a higher score comparing with the original method. We first pass the raw picture through the robust background detection function to get a draft of the salient object. Then we blurred image to eliminate the influence of the color blocks in the draft. By passing the blurred draft image to the edge detection function, we can get a clear outline of the salient object. After combining the blurred and the edge picture, we get a edge strengthened version of salient object draft. Pass this draft to robust background detection again, we get an improved version of salient object.

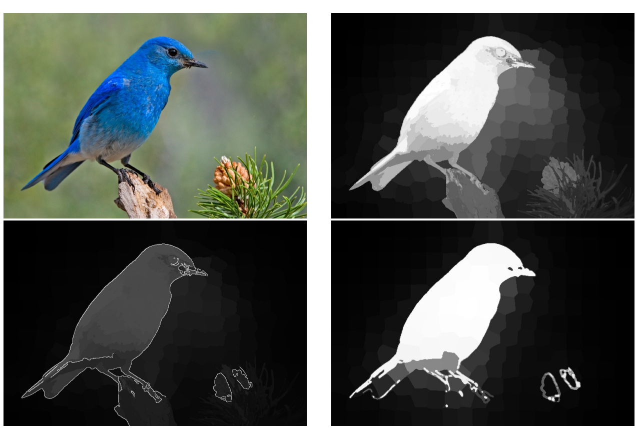
We combine the edge strengthened image and blurred image linearly and in the edge strengthened image, the white pixels are only at the edges. This leads to the result that the object itself looks gray (3rd image). To solve that, we want to further distinguish the object with the background, thus we pick the non-zero pixels and increase the values of the ones that are over the median. Finally we run RBD algorithm again to obtain the final result.

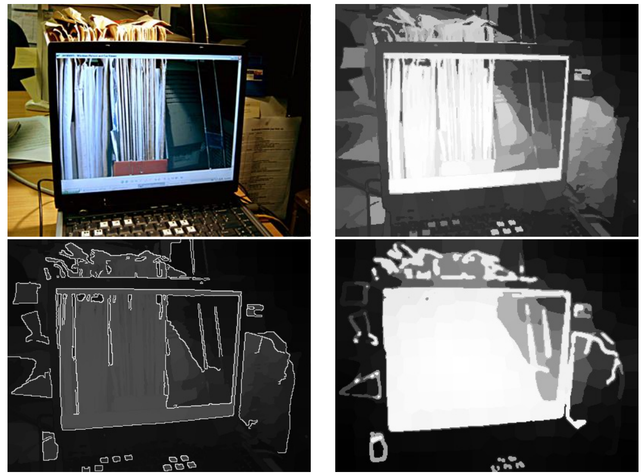
6. Results

From the left to right: **Raw Image, Original RBD Result, Edge detection & blurring, Final improved Result**. We can see that in the original result, it includes the edges in the background, so there is a blurred color block in the result images. By using the edge detection, we successfully remove the polluted background from the result. By passing through the robust detection function again, we get a cleaner version of the salient object. Then we enhance the regions inside the object.









7. Conclusion

We studied the paper *Saliency Optimization from Robust Background Detection*[8] on saliency detection. Based on its weakness of detecting part of the background and incapability of separating the object from the background, we proposed some methods to improve the results.

The basic idea is to run RBD algorithm several times and process the images between different runs. First we run RBD algorithm once. Then we do edge detection on the result to find out the outline of the object. In case the object has different color blocks and the edge detection algorithm will produce edges inside the object, before running the edge detection algorithm, we blurred the image first. After we have the outline of the object, we combine the outline with the blurred image and run the RBD algorithm again. We notice that some area of the object is grey due to linear combination of the two images. Thus we pick the non-zero pixels and increase the values of the ones that are over the median. Finally we re-run RBD algorithm to get the final improvement.

Video Link

`https://drive.google.com/file/d/1-CIj33wY1E48tpUojXZv4P_vhFilqKQ5/view`

References

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P Fua, and S Susstrunk. Global contrast based salient region detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2281, 2015.

[2] M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S. Hu. Global contrast based salient region detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3):569–582, 2015.

[3] Y. Dou, M. Chen, and L. Duan. Salient region detection based on element contrast and boundary prior. *Gaojishu Tongxin/Chinese High Technology Letters*, 25:163–170, 02 2015.

[4] Johnson and D.B. Efficient algorithms for shortest paths in sparse networks. *J.ACM*, 24(1):1–13, 1977.

[5] J. Kim, D. Han, Y. Tai, and J. Kim. Salient region detection via high-dimensional color transform. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 883–890, 2014.

[6] Y. Li, K. Fu, L. Zhou, Y. Qiao, J. Yang, and B. Li. Saliency detection based on extended boundary prior with foci of attention. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2798–2802, 2014.

[7] Vikram Mutneja. Methods of image edge detection: A review. *Journal of Electrical Electronic Systems*, 04, 01 2015.

[8] W. Zhu, S. Liang, Y. Wei, and J. Sun. Saliency optimization from robust background detection. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2814–2821, 2014.