Depth really Matters: Improving Visual Salient Region Detection with Depth report

Paper's Authors

Karthik Desingh, K. Madhava Krishna, Deepu Rajan, and C.V. Jawahar

Advisor

Dr. Maryam Abedi

Student

Mohammad Shahpouri

November 2022

Contents

List of Figures

List of Tables

List of Equations

1 Effect of Depth on Saliency

Competing saliency: The low-contrast object which closer to camera gets more attention when compared to farther regions, Figure 1(a).

Blurred scenes: Not blurred regions are more attentive than the blurred regions as shown in Figure 1(b).

Centre-bias: The low-contrast object placed at the centre of the field of view gets more attention compared to other locations (left and right). Presented as Figure 1(c).

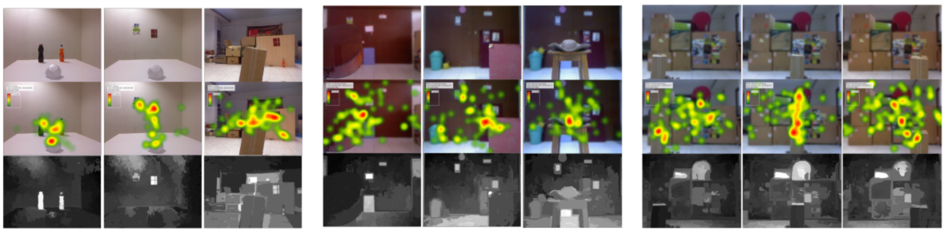
(a) Competing saliency (b) Blurred (c) Centre-bias

Figure 1. (Top) Original images, (Middle) Human eye-fixations shown as a heat map on the count of fixations, (Bottom) Saliency map given by state-of-the-art model RC [1].

2 3D-saliency for Indoor Environment

The region based contrast method from Cheng et al. [1] are adapted to computed contrast. First, histogram H_k for every region R_k are calculated. Contrast score C_k of a region R_k is computed as the sum of the dot products of its histogram with histograms of other regions in the scene. The contrast score is scaled by the depth Z_k of the region R_k which computed by finding the depth of the centroid region. Hence the constrast score becomes:

$$C_k = Z_k \sum_{j \neq k} D_{kj} \tag{1}$$

where D_{kj} is the dot product between histograms H_k and H_j.

To prevent errors in Equation 1 the number of 3D points n_k in region R_k are considered:

$$C_k = \frac{2 Z_k n_k \sum_{j \neq k} D_{kj}}{\sum_j n_j} \tag{2}$$

Saliency of the region R_k becomes $S_k = 1 - C_k / C_{max}$, where C_{max} is the maximum contrast score in the scene for a region.

3 RGBD-Saliency Fusion

To fuse RGB-saliency $S_{rgb}(x, y)$ and 3D-saliency $S_{3D}(x, y)$ values, where (x, y) shows pixels x and y, the SVM regression are adopted.

$$rgbd_i = f(w, f_i, rgb_i, d_i) \tag{3}$$

where w is the weight vector learnt by the SVM model with the help of local feature vector f_i and saliency scores rgb_i, d_i to determine $rgbd_i$ at i^{th} pixel of an image.

Other features used in the fusion process are (along with their feature lengths): *Color Histogram (30)* of region both in terms of RGB and HSV each of 15 bins . *Contour Compactness (1)* is the ratio of the perimeter to the area of the region. *Dimensionality (2)* is the two ratios, minimum dimension by maximum dimension and medium dimension by maximum dimension. *Perspective score (8)* is the ratio of the area projected in the image to the maximum area spread by the region in 3D. *Discontinuities with neighbours (10)* is measure of how much the region is connected with its neighbouring regions. *Size and Location (9)* of the region with respect to the scene gives the range and location of the region in three dimension. *Verticality (20)* is the histogram measure of difference of the normals in the region with respect to the camera pose. They combinely constitute a feature length of 82 along with the RGB and 3D-saliency score.

4 Experimental results

Dataset provided by the University of Washington (UW) [2], Berkeley 3D object dataset [3], and their 33-image dataset are used to evaluate and comparison the method.

4.1 Quantitative result

Table 1. ROC scores of saliency model RC [1] for all three datasets used in this work

Datasets	RGB-RC	D	RGBD-RC	%improvement
Univ of Washington	0.7105	0.7558	0.8053	↑ 9.48%
Berkeley 3D dataset	0.7246	0.7518	0.8157	↑ 9.11%
Thier dataset	0.7287	0.7312	0.8001	↑ 7.14%

4.2 Qualitative result

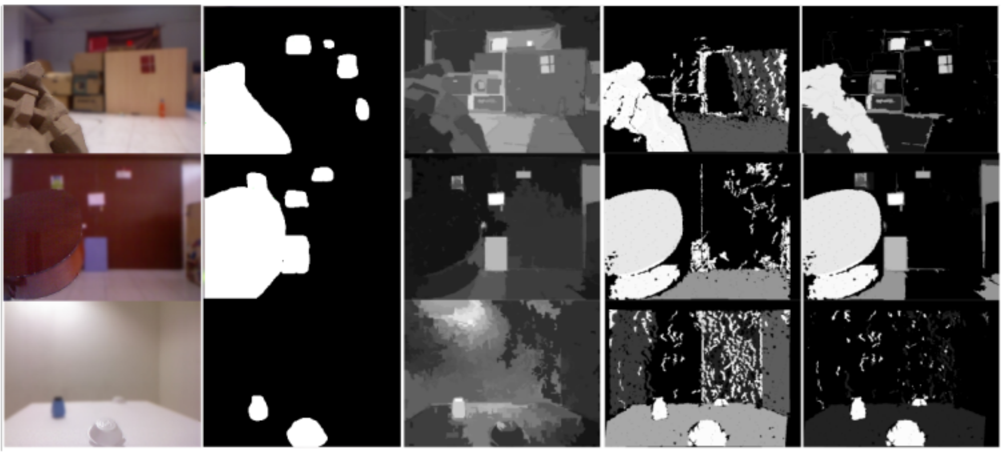

Figure 2. From left (1) Original image and its (2) Human annotated ground truth, (3) RGB saliency map using RC [1], (4) 3D-saliency map and the (5) Fused RGBD-saliency map. RGB-saliency fails to map the objects that are closer, but the fusion of 3D-saliency helps in recovering these objects.

References

[1] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, “Global contrast based salient region detection,” in *CVPR 2011*, pp. 409–416, 2011. ii, iii, 1, 2, 3

[2] K. Lai, L. Bo, X. Ren, and D. Fox, “A large-scale hierarchical multi-view rgb-d object dataset,” in *2011 IEEE International Conference on Robotics and Automation*, pp. 1817–1824, 2011. 2

[3] A. Janoch, S. Karayev, Y. Jia, J. T. Barron, M. Fritz, K. Saenko, and T. Darrell, “A category-level 3-d object dataset: Putting the kinect to work,” in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 1168–1174, 2011. 2