Depth really Matters: Improving Visual Salient Region Detection with Depth

Karthik Desingh[1]
http://researchweb.iiit.ac.in/~karthik.d/
K. Madhava Krishna[1]
http://www.iiit.ac.in/~mkrishna/
Deepu Rajan[2]
http://www.ntu.edu.sg/home/ASDRajan/
C.V. Jawahar[1]
http://www.iiit.ac.in/~jawahar/

[1] IIIT - Hyderabad
Hyderabad, India
[2] Nanyang Technological University
Singapore

Depth information has been shown to affect identification of visually salient regions in images. In this paper, we investigate the role of depth in saliency detection in the presence of (i) competing saliencies due to appearance, (ii) depth-induced blur and (iii) centre-bias. Having established through experiments that depth continues to be a significant contributor to saliency in the presence of these cues, we propose a 3D-saliency formulation that takes into account structural features of objects in an indoor setting to identify regions at salient depth levels. Computed 3D-saliency is used in conjunction with 2D-saliency models through non-linear regression using SVM to improve saliency maps. Experiments on benchmark datasets containing depth information show that the proposed fusion of 3D-saliency with 2D-saliency models results in an average improvement in ROC scores of about 9% over state-of-the-art 2D saliency models.

The main contributions of this paper are: (i) The development of a 3D-saliency model that integrates depth and geometric features of object surfaces in indoor scenes. (ii) Fusion of appearance (RGB) saliency with depth saliency through non-linear regression using SVM. (iii) Experiments to support the hypothesis that depth improves saliency detection in the presence of blur and centre-bias. The effectiveness of the 3D-saliency model and its fusion with RGB-saliency is illustrated through experiments on two benchmark datasets that contain depth information. Current state-of-the-art saliency detection algorithms perform poorly on these datasets that depict indoor scenes due to the presence of competing saliencies in the form of color contrast. For example in Fig. 1, saliency maps of [1] is shown for different scenes, along with its human eye fixations and our proposed saliency map after fusion. It is seen from the first scene of Fig. 1, that illumination plays spoiler role in RGB-saliency map. In second scene of Fig. 1, the RGB-saliency is focused on the cap though multiple salient objects are present in the scene. Last scene at the bottom of Fig. 1, shows the limitation of the RGB-saliency when the object is similar in appearance with the background.

Effect of depth on Saliency: In [4], it is shown that depth is an important cue for saliency. In this paper we go further and verify if the depth alone influences the saliency. Different scenes were captured for experimentation using Kinect sensor. Observations resulted out of these experiments are (i) Humans fixate on the objects at closer depth, in the presence of visually competing salient objects in the background, (ii) Early attention happens on the objects at closer depth, (iii) Effective fixations are high at the low contrast foreground compared to the high contrast objects in the background which are blurred, (iv) Low contrast object placed at the center of the field of view, gets more attention compared to other locations. As a result of all these observations, we develop a 3D-saliency that captures the depth information of the regions in the scene.

3D-Saliency: We adapt the region based contrast method from Cheng *et al*. [1] in computing contrast strengths for the segmented 3D surfaces or regions. Each segmented region is assigned a contrast score using surface normals as the feature. Structure of the surface can be described based on the distribution of normals in the region. We compute a histogram of angular distances formed by every pair of normals in the region. Every region R_k is associated with a histogram H_k. 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. Since the depth of the region is influencing the visual attention, the contrast score is scaled by a value Z_k, which is the depth of the region R_k from the sensor. In order to define the saliency, sizes of the regions i.e. the number of the points in the region, have to be considered. We find the ratio of the region dimension to the half of the scene dimension. Considering n_k as the number of 3D points in the region R_k, the constrast score becomes

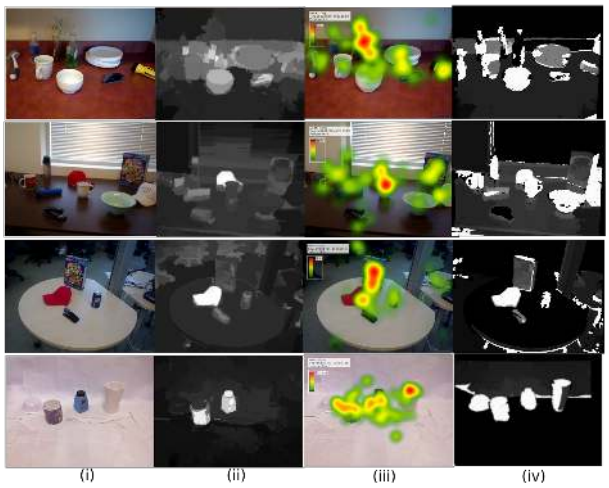

Figure 1: Four different scenes and their saliency maps; For each scene from top left (i) Original Image, (ii) RGB-Saliency map using RC [1], (iii) Human fixations from eye-tracker and (iv) Fused RGBD-saliency map

$$C_k = \frac{2Z_k n_k \sum_{j \neq k} D_{kj}}{\sum_j n_j} \quad (1)$$

where D_{kj} is the dot product between histograms H_k and H_j. The region with less C score is considered to be the one that is unique in the 3D scene. Hence 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. With the obtained 3D-saliency map, we fuse saliency maps given by the state-of-the-art algorithms to obtain the RGBD-saliency map.

Fusion process: The obtained 3D-saliency map is fused with the saliency maps of existing visual saliency models to create RGBD-saliency. We perform this operation using SVM regression model trained on set of sample images. This fusion process also includes additional local features of the 3D region in the training. They are *Color histogram*, *Contour compactness*, *Dimensionality*, *Perspective score*, *Discontinuities with neighbours*, *Size and location*, and *Verticality*. Inclusion of these features improves the performance by more than 5%. Results are quantified across three datasets, University of Washington RGB-D dataset [3], Berkely 3D dataset [2] and our own dataset.

Our conclusion is that the depth is an important cue for the detection of salient region in a scene. With a novel formulation we propose 3D-saliency which when fused with saliency of the existing models enhances their performance.

[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.

[2] A. Janoch, S. Karayev, Y. Jia, J.T. Barron, M. Fritz, K. Saenko, and T. Darrell. A Category-Level 3D Object Dataset: Putting the Kinect to Work. In *ICCV Workshop*. 2011.

[3] K. Lai, L. Bo, X. Ren, and D. Fox. A Large-Scale Hierarchical Multi-View RGB-D Object Dataset. In *ICRA*. 2011.

[4] C. Lang, T.V. Nguyen, H. Katti, K. Yadati, M. Kankanhalli, and S. Yan. Depth Matters: Influence of Depth Cues on Visual Saliency. In *ECCV*. 2012.