# DEPTH SALIENCY BASED ON ANISOTROPIC CENTER-SURROUND DIFFERENCE

*Ran Ju, Ling Ge, Wenjing Geng, Tongwei Ren and Gangshan Wu*

State Key Laboratory for Novel Software Technology
Nanjing University, China
juran@smail.nju.edu.cn, gelingnju@gmail.com, wjgeng@smail.nju.edu.cn,
rentw@nju.edu.cn, gswu@nju.edu.cn

## ABSTRACT

Most previous works on saliency detection are dedicated to 2D images. Recently it has been shown that 3D visual information supplies a powerful cue for saliency analysis. In this paper, we propose a novel saliency method that works on depth images based on anisotropic center-surround difference. Instead of depending on absolute depth, we measure the saliency of a point by how much it outstands from surroundings, which takes the global depth structure into consideration. Besides, two common priors based on depth and location are used for refinement. The proposed method works within a complexity of $O(N)$ and the evaluation on a dataset of over 1000 stereo images shows that our method outperforms state-of-the-art.

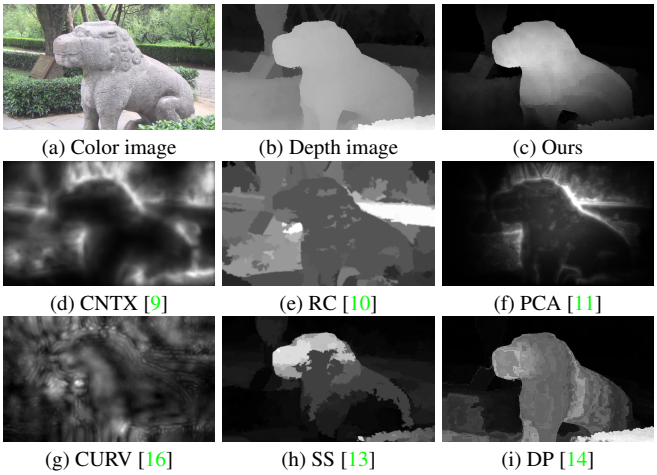
***Index Terms*—** Saliency detection, depth image

## 1. INTRODUCTION

Saliency detection [1] is also regarded as visual attention for human. The activity is a complex process including visual information gathering and filtering, with its aim to find the most conspicuous regions rapidly from sight. By only selecting the salient subset for further processing, the complexity of higher visual analysis can be reduced significantly. Many applications benefit from saliency analysis, e.g. object segmentation [2], image classification [3, 4], image/video retargeting [5, 6], compression [7] and quality assessment [8].

Computational saliency model [1] performs a feature integration process similar to human activity, which first extracts features from input visual information and then integrates them into a saliency map. Both visual information gathering and feature extraction contribute largely to saliency detection. For convenience, most existing works [9, 10, 11] take a 2D color image as input, which turns out to be insufficient in certain cases. We give a few examples of saliency detection results in Fig. 1 (d)-(f).

This work is supported by the National Science Foundation of China under Grant No.61321491 and No.61202320, Research Project of Excellent State Key Laboratory (No.61223003), Natural Science Foundation of Jiangsu Province (No.BK2012304) and National Special Fund (No.2011ZX05035-004-004HZ).



(a) Color image (b) Depth image (c) Ours
(d) CNTX [9] (e) RC [10] (f) PCA [11]
(g) CURV [16] (h) SS [13] (i) DP [14]

**Fig. 1**. Saliency maps generated by different methods. This figure illustrates that depth information supplies a powerful cue for saliency prediction.

A few studies try to investigate the effect of scene depth for saliency [12]. And recently it has been shown that depth perception has a strong impact in visual attention [13, 14]. As depth image differs a lot from color image, feature extraction turns out to be a key problem in depth saliency analysis. Early visual features like depth [15], gradient and curvature [16] easily lead to erroneous detection for the lack of global consideration, as shown in Fig. 1 (b) and (g). Stereo saliency [13] prefers unique and nearer regions. However, the basic assumption that salient regions have zero disparities or high contrasts to nearby regions turns to be limited and may easily miss inner parts of salient objects, as shown in Fig. 1 (h). Lang et al. [14] model saliency as the the conditional probability given depth and depth range. The limitation is that it only cares about absolute depth while missing the global depth structure information, as shown in Fig. 1 (i).

In this paper we present a novel saliency detection method based on following considerations:

- The depth prior that nearer regions appear more salient is effective but may be easily puzzled by nearer backgrounds. Besides, two regions with the same depth but

different surroundings should be differentiated.

- Salient objects tend to outstand from surrounding backgrounds. The surroundings should be globally considered because the inner part of an object tends to be flat but the entire object may be protruding. This assumption is more effective than depth prior because it prefers relative depth to absolute depth with considering of scene structures, which can be seen from the stereo and human eye fixation dataset [13, 14].
- Center regions are more salient than peripheral due to the common photography tendency, which has been shown in previous works [11].
- The computation for saliency should be efficient for higher level visual processing tasks as stated in [1].

Our method arises from the above considerations. We define the depth saliency of a point as how much it outstands from surroundings, which is measured using an anisotropic center-surround operator. Besides, we employ the depth and center priors for refinement. Considering of efficiency and robustness we perform saliency detection on the superpixel granularity. Our method is $O(N)$ complex where $N$ is the number of image pixels. For evaluation we build a dataset that contains over 1000 stereo images with salient object masks. The results show that our method can outperform state-of-the-art on detecting salient regions.

## 2. APPROACH

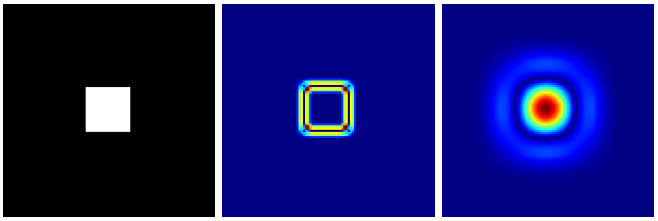
### 2.1. Depth Acquisition

We first consider how to acquire depth information. There are a lot of devices to capture depth like Time of Flight (ToF) camera, laser range scanner, structured light scanner etc. In this work we choose to recover depth maps from stereo images because they are easy to capture and popular in daily life. And hence we can easily collect data for evaluation from resources like image website, daily life photographing and 3D movie snapshots. The depth image is generated using Sun's optical flow method [17] for its accuracy and robustness. An example is shown in Fig. 1 (b) where nearer pixels appear brighter and vice versa.

### 2.2. Anisotropic Center-Surround Difference

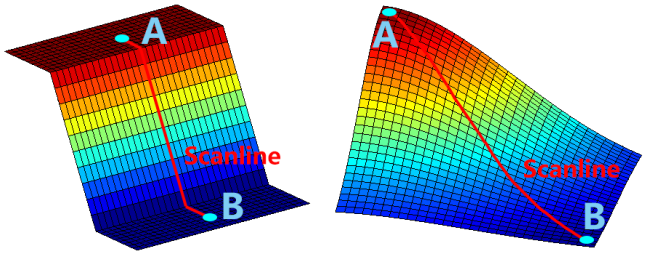
According to our basic assumption that visual attention is more easily to be paid to regions outstanding from surroundings, we search for a depth feature for measurement. A simple description is the center-surround operator such as Difference of Gaussian (DoG). The limitation is that DoG misses global information on a fine scale and ignores details on a large scale as illustrated in Fig. 2.

To overcome this problem, we propose to perform an anisotropic scan along multiple directions. In each scanline, we assume the pixel with the minimum depth value as background and calculate the difference between the center



(a) Input (b) $\sigma = 5$ (c) $\sigma = 25$

**Fig. 2**. Limitation of DoG. (a) Input image. The white box appears conspicuous. (b) Only edge areas of the white box are detected. (c) The edge details are missed.




(a) Sharp edge (b) Smooth edge

**Fig. 3**. Sharp and smooth edges comparison.

pixel and background. The depth change in between is not considered because we believe whether a sharp nor smooth edge would change the saliency of the center pixel. An example is shown in Fig. 3. The saliency values of point $A$ in the sharp and smooth cases are assumed the same due to the equal magnitude they stand out from the lower point $B$. The depth change only affects the saliency of the boundary areas.

Compared to DoG, the proposed operator performs an anisotropic center-surround difference (ACSD) measure. Obviously the operator is easily to be interfered by noise. So we first perform a Gaussian smoothing with $\sigma_s = 7$ on the depth image. Besides, considering that distant pixels are less significant, we set a maximum scan length $L$ for each scanline. In our experiments we set $L$ as a third of the diagonal length. The ACSD is summed over eight scanning directions as shown in Fig. 4. Now we give the mathematical description of our anisotropic center-surround difference measure:

$$D^i_{acsd}(p) = d(p) - \min(d^i_k), k \in [1, L] \quad (1)$$

$$D_{acsd}(p) = \sum_{i \in [1,8]} D^i_{acsd}(p) \quad (2)$$

where $D^i_{acsd}(p)$ indicates the ACSD value of pixel $p$ along the scanline $i$. $d(p)$ is the depth value of pixel $p$. $k$ is the index of the pixels along the scan path $i$ and limited to $L$. $D_{acsd}(p)$ is the ACSD of pixel $p$ which sums the ACSDs in eight directions. In Fig. 4 we show the pixels with minimum depth values in each scanline. In this example the center point gets a high saliency as it appears outstanding in all the scanning directions. What we concerns more is the ground that extends from far to near. Obviously the distant background gets very

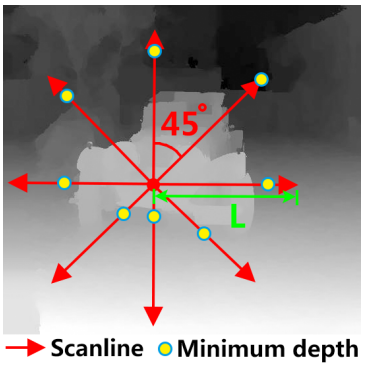



**Fig. 4**. Example of the ACSD operator.

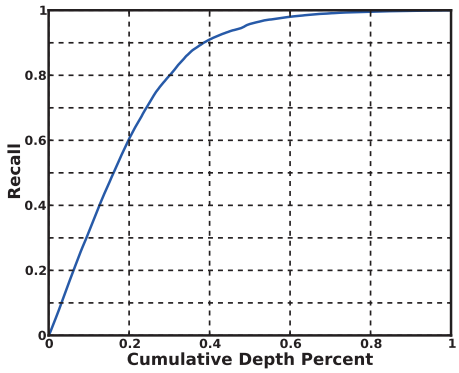



**Fig. 5**. Recall versus cumulative depth percent curve.

low ACSD value and thus looks inapparent. The nearer part of the ground, which is located at the bottom of the image, has a definitely high depth value. However, it shows less conspicuity because it gets high ACSD values only in the upper three directions. In the horizontal and lower directions it is inhibited effectively.

### 2.3. Saliency Computation and Refinement

To deal with noises and errors in depth images, the saliency detection is performed on the granularity of superpixels, which is generated using SLIC [18] on the color images. The number of superpixels is set as the length of diagonal in pixels. Then we compute the ACSD for each superpixel at the centroid. The depth value of the centroid is calculated as the mean depth value over the superpixel. Then the saliency is rescaled to $[0, 255]$ and assigned to each pixel to form an initial saliency map $S_{acsd}$.

Next we refine the initial result using two common priors. First is that regions nearer to viewers appear more salient. We may binarize the depth image using a threshold varying from near to far (depth value ranges from 255 to 0) to calculate the recall versus cumulative depth percent curve. As illustrated in Fig. 5, the top 50% nearer pixels gives a 95.78% recall rate of the salient region. And thus we remain the saliency of top 50% close pixels unchanged and add a linear weighting $d(p)/d_{50}$ to the remaining pixels, where $d_{50}$ is the top 50% threshold. The second prior is that salient objects tend to locate at the center [11, 19]. Similar to [11], we add a 2D Gaussian $G(x, y, \sigma_x, \sigma_y)$ centered at the image with $\sigma_x$ and $\sigma_y$ equal to half of the width and height of the image.

### 2.4. Complexity Analysis

SLIC is $O(N)$ complex and generate $O(L)$ superpixls where $L$ is a third of the diagonal length. ACSD is also $O(L)$ complex as the scan length is limited to $L$. And thus the $S_{acsd}$ computation is $O(L^2)$ complex. Suppose the aspect ratio of the image is $r$, we can get $L^2 = \frac{r^2+1}{r}N$. The complexity of depth and center prior based refinement is also $O(N)$. According to the above we conclude that our saliency method works within an $O(N)$ complexity.

## 3. EXPERIMENTS AND ANALYSIS

### 3.1. Dataset and Experiment Settings

We collect stereo images from Internet, 3D movies and photographs taken by a Fuji W3 stereo camera. As the labeling results on 2D images may be a little different from that in real 3D environments, we perform mask labeling in a 3D display environment by using *Nvidia 3D Vision*. We first collect more than 20000 stereo images totally. Next, following the procedure in [20], we selected 5913 images each of which contains a salient object of moderate size. After that, four volunteers are invited to label the salient object masks. At last, 1382 high quality and consistently labeled images are selected for evaluation.

We employ the widely used precision-recall curve to evaluate the performance of our method. Specifically, we obtain a binary image from the saliency map using a gradually increasing threshold from 0 to 255 and compare with the groundtruth salient object mask to get the precision and recall. We choose three state-of-the-art methods work on color images namely CNTX [9], RC [10], PCA [11] and three depth saliency methods namely CURV [16], SS [13], DP [14] for comparison. Besides, the depth images are directly treated as saliency maps (named DEPTH) to evaluate the depth prior.

### 3.2. Results and Discussion

We show a few saliency maps generated by different methods in Fig. 6. The precision-recall curves are given in Fig. 7. Almost all of the depth image based methods perform better than the color image based methods. An exception is CURV, which is based on depth curvature. The precision decreases quickly as recall increases because CURV detects local regions instead of entire objects. And this suggests that curvature is not proper for salient object detection task in depth images. The maximum precision of the three color image based methods would not exceed 0.6, while DEPTH reaches a precision more than 0.7. An explanation is that salient objects may look inapparent in color or context but conspicuous in 3D perception. This suggests that 2D color information is insufficient for saliency detection and 3D


**Fig. 6**. Saliency comparisons of different methods. The first column shows the left views of the stereo images. The second and third column shows the depth images and ground truth salient object masks repectively. The next three columns are the saliency results of color image based methods. The last four columns show the results of depth saliency methods.

depth cues may supply a more powerful prediction for visual attention in certain cases. Similar results and conclusions are stated in [13, 14]. An interesting phenomenon is that DEPTH keeps almost a constant precision rate of 0.7 at the recall rates from 0.1 to 0.8. This corresponds to the top 5% to 30% close pixels in Fig. 5. That is to say, in this range each pixel has a 70% probability to appear salient. The DP method which leverages prior probabilities shows no obvious improvement to DEPTH. This can be explained that relative depth contributes more to saliency than absolute depth values. SS performs the most close to our method as it takes relative depth into consideration. As stated in Section 1, the limitation is that its preference to unique regions may miss inner regions of objects, which can be seen in Fig. 6.

We implement our method in C++ and test on a machine with a 3.4GHz Intel i7-4770 CPU and 16GB memory. Typically for a $1280 \times 720$ image, the running time is 0.718s. Specifically, the superpixel segmentation takes 0.656s and saliency computation takes 0.062s.

## 4. CONCLUSION

We proposed a salient object detection method that works on depth images using anisotropic center-surround difference. The method is based on a simple but effective assumption that salient objects tend to stand out from surrounding background. Furthermore, two priors based on depth and location are included for saliency refinement. Our method is fast and works within a linear complexity. The experiments demonstrated that our method can be used for rapid and accurate salient object detection task.

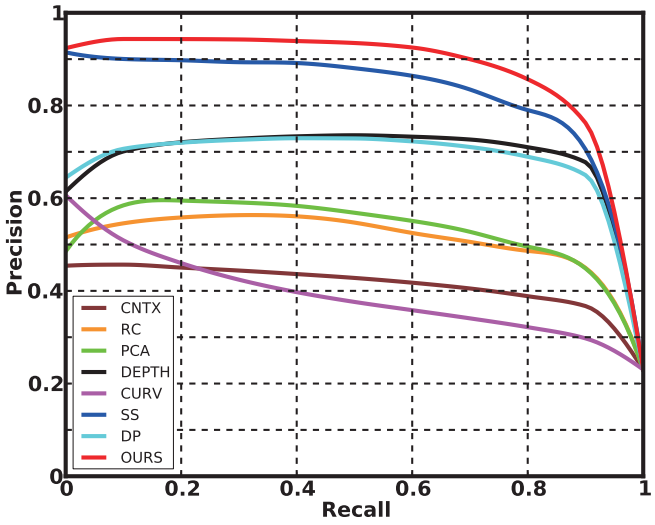



**Fig. 7**. Precision-recall curves of different methods.

## 5. REFERENCES

[1] Laurent Itti, Christof Koch, and Ernst Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[2] Junwei Han, King Ngi Ngan, Mingjing Li, and Hong-Jiang Zhang, "Unsupervised extraction of visual attention objects in color images," *IEEE Transactions*

*on Circuits and Systems for Video Technology*, vol. 16, no. 1, pp. 141–145, 2006.

[3] Gaurav Sharma, Frédéric Jurie, and Cordelia Schmid, "Discriminative spatial saliency for image classification," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 3506–3513.

[4] Lai-Kuan Wong and Kok-Lim Low, "Saliency-enhanced image aesthetics class prediction," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*. IEEE, 2009, pp. 997–1000.

[5] Matthias Grundmann, Vivek Kwatra, Mei Han, and Irfan Essa, "Discontinuous seam-carving for video retargeting," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 569–576.

[6] Daniel Domingues, Alexandre Alahi, and Pierre Vandergheynst, "Stream carving: an adaptive seam carving algorithm," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE, 2010, pp. 901–904.

[7] Laurent Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1304–1318, 2004.

[8] Xin Feng, Tao Liu, Dan Yang, and Yao Wang, "Saliency based objective quality assessment of decoded video affected by packet losses," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE, 2008, pp. 2560–2563.

[9] Stas Goferman, Lihi Zelnik-Manor, and Ayellet Tal, "Context-aware saliency detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1915–1926, 2012.

[10] Ming-Ming Cheng, Guo-Xin Zhang, Niloy J Mitra, Xiaolei Huang, and Shi-Min Hu, "Global contrast based salient region detection," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2011, pp. 409–416.

[11] R. Margolin, A. Tal, and L. Zelnik-Manor, "What makes a patch distinct?," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2013, pp. 1139–1146.

[12] Nabil Ouerhani and H Hugli, "Computing visual attention from scene depth," in *International Conference on Pattern Recognition*. IEEE, 2000, vol. 1, pp. 375–378.

[13] Yuzhen Niu, Yujie Geng, Xueqing Li, and Feng Liu, "Leveraging stereopsis for saliency analysis," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 454–461.

[14] Congyan Lang, Tam V Nguyen, Harish Katti, Karthik Yadati, Mohan Kankanhalli, and Shuicheng Yan, "Depth matters: Influence of depth cues on visual saliency," in *European Conference on Computer Vision*, pp. 101–115. Springer, 2012.

[15] Sungmoon Jeong, Sang-Woo Ban, and Minho Lee, "Stereo saliency map considering affective factors and selective motion analysis in a dynamic environment," *Neural Networks*, vol. 21, no. 10, pp. 1420–1430, 2008.

[16] Chang Ha Lee, Amitabh Varshney, and David W Jacobs, "Mesh saliency," in *ACM Transactions on Graphics*. ACM, 2005, vol. 24, pp. 659–666.

[17] Deqing Sun, Stefan Roth, and Michael J Black, "Secrets of optical flow estimation and their principles," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 2432–2439.

[18] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[19] Tongwei Ren, Ran Ju, Yan Liu, and Gangshan Wu, "How important is location in saliency detection?," in *ACM International Conference on Internet Multimedia Computing and Service*. IEEE, 2014.

[20] Tie Liu, Zejian Yuan, Jian Sun, Jingdong Wang, Nanning Zheng, Xiaoou Tang, and Heung-Yeung Shum, "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353–367, 2011.