Exploiting Global Priors for RGB-D Saliency Detection

Jianqiang Ren, Xiaojin Gong, Lu Yu
Zhejiang University
{rjq, gongxj, yul}@zju.edu.cn

Wenhui Zhou
Hangzhou Dianzi University
zhouwenhui@hdu.edu.cn

Michael Ying Yang
TU Dresden
Ying.Yang1@tu-dresden.de

Abstract

Inspired by the effectiveness of global priors for 2D saliency analysis, this paper aims to explore those particular to RGB-D data. To this end, we propose two priors, which are the normalized depth prior and the global-context surface orientation prior, and formulate them in the forms simple for computation. A two-stage RGB-D salient object detection framework is presented. It first integrates the region contrast, together with the background, depth, and orientation priors to achieve a saliency map. Then, a saliency restoration scheme is proposed, which integrates the PageRank algorithm for sampling high confident regions and recovers saliency for those ambiguous. The saliency map is thus reconstructed and refined globally. We conduct comparative experiments on two publicly available RGB-D datasets. Experimental results show that our approach consistently outperforms other state-of-the-art algorithms on both datasets.

1. Introduction

Although saliency detection has been extensively studied for decades, most of the previous works focus on 2D image analysis. Recently, researchers began to integrate multi-modality imaging data to improve performance. For instance, flash and no-flash image pairs [11], light field [19], stereopsis [21], or RGB-D images [16, 8, 22, 6] are investigated respectively. Among them, the use of RGB-D images for salient object detection is attracting more and more interest, partially due to the advent of robust ranging sensors such as the Microsoft Kinect and Velodyne Lidars, but also because of its importance on navigation and manipulative tasks in robotics.

In both cognitive psychology [28, 9] and computer vision [12] communities, it is well known that 3D spatial information is an important cue for human visual attention. Color images augmented with depth maps provide us with scene layout, shape, surface orientation and other 3D cues, making it possible to pop out salient objects even if foreground and background are similar in appearance. Several prior works [16, 8, 6, 22, 14] have tried to integrate depth with color for enhancing saliency detection performance. Most of these methods incorporate depth cues within the center-surround contrast measurement framework. Global priors induced by depth have not been well developed yet, except the work in [16].

Contrarily, global priors have exhibited their high effectiveness in 2D saliency detection. For instance, center-bias [27] is a prior that has been extensively used to boost performance. Boundary prior [30, 32, 33] and semantic prior [26] are also exploited in some recent approaches, by which state-of-the-art results are obtained, demonstrating the advantage of using the global knowledge. Inspired by these achievements, this work aims to explore particular priors that are introduced by the depth information, meanwhile investigating the effectiveness of 2D priors for RGB-D saliency analysis.

The contribution of this paper lies in three aspects. First, we propose two priors, which are normalized depth prior and global-context surface orientation prior, for RGB-D saliency detection. The former is formulated in a simple form, based on an observation from saliency distributions, to highlight near objects while degenerate saliency for objects further away. The orientation prior is estimated with respect to the principal axis, aiming to degrade saliency for severely slanted surfaces such as the ground plane or ceilings. Combining these two intuitive but highly complementary priors with the traditional region contrast framework, we obtain high detection performance. Second, we investigate the effectiveness of 2D background prior for RGB-D images and find out that it is of a marginal contribution to boost the performance. The main reason is that, with depth information, objects are easier to be popped out no matter if they are connected to boundaries or not. Last but not least, we propose a saliency restoration scheme to refine saliency detection results. It integrates the PageRank algorithm [4] to sample high confident regions and reconstruct saliency for the ambiguous regions via globally optimizing a Markov Random Field. The effectiveness of our approach has been validated on two publicly available RGB-D datasets.

2. Related Work

2D Saliency: 2D saliency detection is originally predicting eye fixations on images, and later is extended to identify salient object regions. Over the past decades, a great number of methods have been developed, from either biological or top-down and bottom-up computational standpoints. The top-down computational models are driven by some specific tasks while the bottom-up models, which is the focus of most recent researches, rely on input images [3].

Many existing bottom-up approaches exploit contrast priors [30] for saliency detection. The center-surround contrast of image elements, either pixels or superpixels, is computed with respect to local neighborhoods [15, 10, 31] or the entire image [1, 5]. Diverse features such as color difference and edge orientations [15], and features at multiple scales [31] or in frequency domain [18] are investigated for contrast measurement. With an estimated saliency map, some recent work further applies a Grabcut-like scheme [25], a Markov Random Field model [31, 6], or other optimization framework [33] to globally optimize the final result, boosting the performance in salient object detection tasks.

Besides the commonly used center-surround contrast, other priors are also incorporated in recent approaches. For example, center bias [27], which refers to the tendency for humans to look towards the image center, is often used at the final step for saliency detection. Wei *et al*. [30] propose a background prior to eliminate salient regions on image boundaries, which demonstrates high effectiveness [30, 32, 33]. Moreover, semantic prior [26], assuming that people pay more attention to certain semantic objects like faces, is also exploited. The incorporation of these complementary priors improves the saliency detection performance further. However, even if many methods have achieved high performance on different datasets, they are prone to fail when foreground and background are similar in appearance, due to the lack of 3D information.

3D Saliency: Existing 3D saliency detection researches are conducted on different data modalities, including 3D point clouds [23], stereopsis [21, 14], and RGB-D images [16, 8, 22, 6]. Analogous to 2D saliency analysis, local center-surround operators are applied to detect salient 3D meshes [17] and points [23], based on mean curvatures, surface height, relative surface orientation and other 3D cues. Niu *et al*. [21] employ a global disparity contrast to leverage stereopsis for saliency detection. Ju *et al*. [14] use anisotropic center-surround difference to measure saliency on depth images. For RGB-D images, one research line treats the depth map independently to generate a depth-induced saliency map and then combine it with a color-produced saliency map via some heuristic or machine learning approaches [8]. The other line combines 3D cues derived from the depth map with color information at each step of saliency detection [22]. When examining depth information, an observed prior is that objects closest to humans always attract the most attention [29], while attention attenuates greatly for further objects. Such depth prior is learnt via Gaussian mixture models in [16] to predict eye fixation on RGB-D images. Other priors, such as the comfort zone bias [21] which is specific to the stereopsis domain, are also exploited to improve the performance. However, in contrast to 2D saliency analysis, the priors induced by 3D information have not been well exploited in the existing works. Moreover, even the effectiveness of the 2D priors, background prior for example, has not been evaluated when applied to 3D saliency detection.

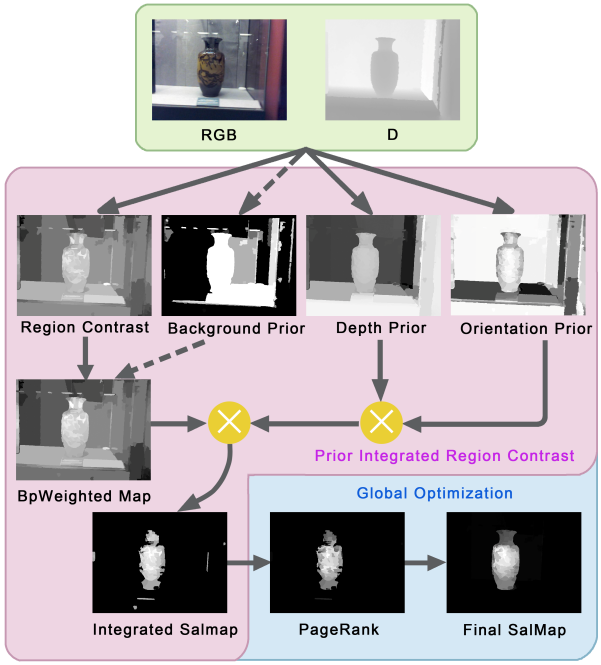


Figure 1: An overview of the proposed framework. (The dash line along the background prior (BP) means that the framework can choose to either incorporate BP or not.)

3. Proposed Method

This paper presents a two-stage scheme for RGB-D saliency detection. The first stage generates a saliency map based on the combination of region contrast with global priors. Specifically, when an RGB image and an aligned depth map are given, we first convert the RGB image to the CIE LAB color space and oversegment it into superpixels using Mean Shift [7], SLIC [2] or other methods. The contrast of each superpixel is evaluated with respect to all other superpixels. Three global priors that are, respectively, the background, depth, and surface orientation priors, are exploited and incorporated together with the region contrast to

produce a saliency map. At the second stage, a PageRank-based sampling scheme and a MRF-based saliency restoration method is applied to refine the final result globally. Figure 1 illustrates the overall framework.

3.1. Prior Integrated Region Contrast

3.1.1 Region Contrast

Our region contrast is designed in a form similar to others [5], but incorporating the depth value at the same time. For each superpixel, or we say a region r_i, its saliency value is measured with respect to its color and depth contrast to all other superpixels

$$S_{rc}(r_i) = \sum_{j \neq i} A(r_j) C(r_i, r_j), \tag{1}$$

where $i, j \in \{1, ..., n\}$ are the indexes of regions and n is the total number, $A(r_j)$ indicates the area of region r_j, and

$$C(r_i, r_j) = \exp\left(-\frac{||\mathbf{x}_i - \mathbf{x}_j||_2^2}{2\sigma_{\mathbf{x}}^2}\right) ||\mathbf{f}_i - \mathbf{f}_j||_2, \tag{2}$$

in which $\mathbf{f} = [l, a, b, d]$ stands for the color vector augmented with depth, taking the average value within a superpixel and $\mathbf{x}$ denotes the superpixel's centroid. $\sigma_{\mathbf{x}}$ is the standard deviation of the distance between two centroids.

3.1.2 Background Prior

Background prior has been exploited in recent 2D [30, 13, 32, 33] and 3D [22] saliency detection approaches. It relies on two observations. One assumes that boundary regions are mostly background, because photographers seldom crop objects of interest along the view frame. The other is that background regions are more of high connectivity. Therefore, Wei *et al*. [30] proposed a geodesic saliency to measure background regions of shorter paths. Jiang *et al*. [13] designed a backgroundness descriptor to determine the background degree of a region via a supervised learning approach. Yang *et al*. [32] used appearance connectivity to boundaries for saliency evaluation. Peng *et al*. [22] simply chose four image corners as pseudo-background. These approaches have demonstrated their effectiveness. However, they are mostly heuristic and fragile when objects touch the boundary even slightly.

In this work, we adopt boundary connectivity [33] to generate the background prior map, which is defined as

$$S_{bp}(r_i) = 1 - \exp\left(-\frac{L(r_i)^2}{2\sigma_b^2 A(r_i)}\right), \tag{3}$$

in which $L(r_i)$ refers to the length along the boundary for region r_i and σ_b is a weighting factor for boundary connectivity. This background measure is robust to the slightly touched cases. As shown in Figure 2d, the approach eliminates most boundary regions effectively. However, in many cases, it still keeps a large portion of an image as saliency. Moreover, it inevitably fails when an object heavily connects to the image boundary, as the last example shown in Figure 2d.

3.1.3 Depth Prior

A depth prior commonly observed in both vision perception and cognitive psychology is that closest objects attract the most attention. In addition, Lang *et al*. [16] pointed out that the relation between depth and saliency is non-linear. They thereby employed a Gaussian Mixture Model to learn the depth prior, which characterizes saliency versus depth within different depth of fields (DOF).

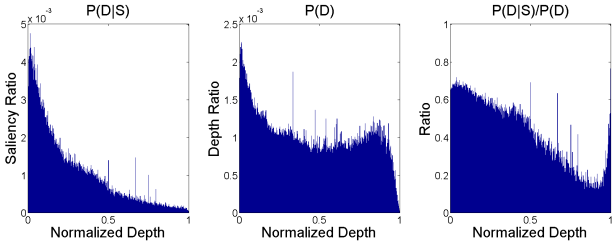
Considering that salient objects may occur at quite different positions when the scenes have different DOFs, we therefore rescale the absolute depth values within each image into the range $[0, 1]$ to get rid of the influence of DOFs. Figure 3a and Figure 3b illustrate the depth distribution of salient objects and the distribution for all scenes, respectively, corresponding to the probabilities $P(D|S)$ and $P(D)$. Given a depth value D, we maximize the posterior probability $P(S|D)$ that is

$$P(S|D) \propto \frac{P(D|S)}{P(D)}. \tag{4}$$

Figure 3c demonstrates the distribution of $\frac{P(D|S)}{P(D)}$. Based on that, we simply use the following form to estimate the depth prior map

$$S_{dp}(r_i) = \sqrt{1 - d(r_i)}, \tag{5}$$

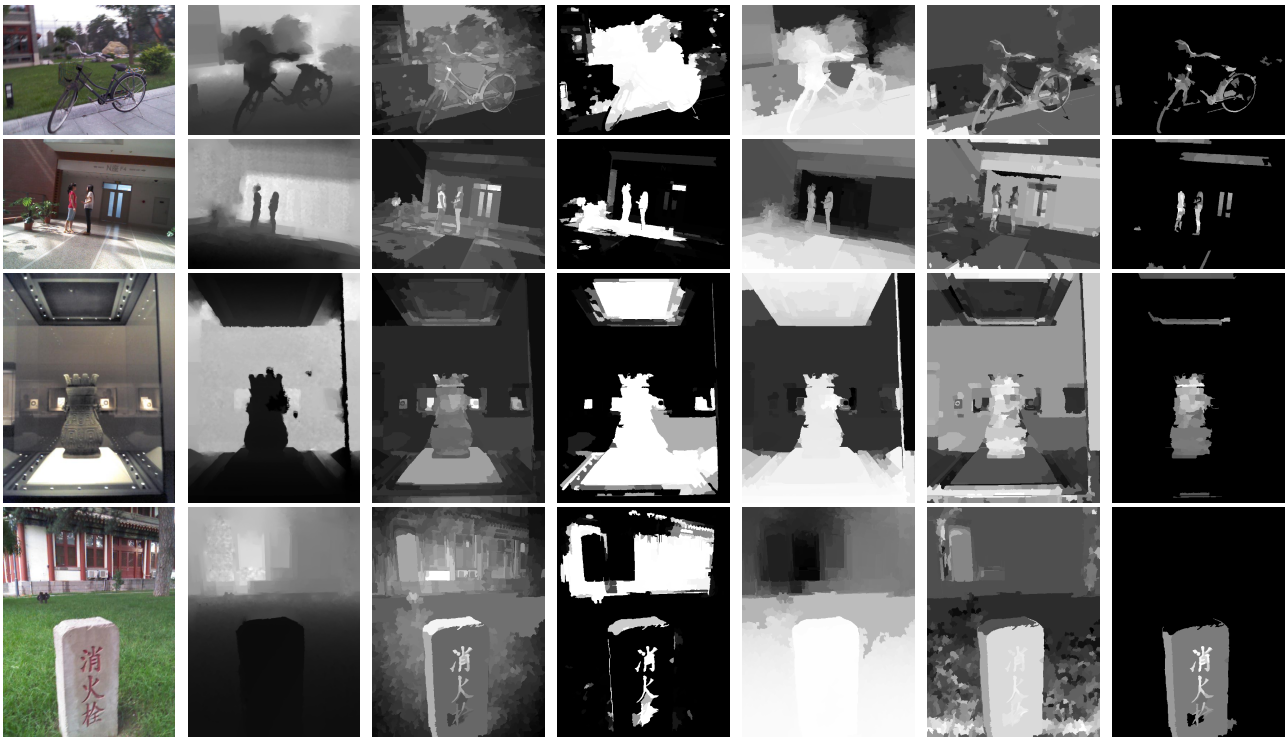
where $d(r_i)$ is the mean depth value of region r_i. This formulation roughly fits the posterior distribution, while placing a small bias on closer distance.



(a) (b) (c)

Figure 3: The depth distributions obtained in the RGB-D dataset [22].

Figure 2e presents some depth prior maps produced according to Eq.(5), in which closer objects are popped out



(a) Image (b) Depth (c) Region Contrast (d) Background Prior (e) Depth Prior (f) Orientation Prior (g) Integration

Figure 2: Typical examples of saliency detection results generated at each step.

from further backgrounds notably. However, when an object is placed on a support plane such as the ground, the closer portion of the plane is highlighted as well.

3.1.4 Surface Orientation Prior

Although surface orientation is considered to be an important 3D cue in several RGB-D approaches, most of them [22, 6, 23] measure the relative orientation between each two superpixels for region contrast evaluation. In this work, we propose a global-context orientation prior. It is based on an intuitive observation that photographers like to choose a viewpoint fronto-parallel to objects of interest, or we say that the direction perpendicular to the principal axis receives the most attention. Denoting the principal axis by $\mathbf{z}$, we estimate the orientation prior by

$$S_{op}(r_i) = \langle \mathbf{z}, \mathbf{n}(r_i) \rangle \,, \tag{6}$$

where $\mathbf{n}(r_i)$ is the unit surface normal of region r_i and $\langle \cdot, \cdot \rangle$ denotes the inner product.

Figure 2f demonstrates four produced orientation prior maps. It shows that the saliency on the ground plane, ceiling, or sidewalls is greatly degenerated. We also observe that surface orientation prior and depth prior are highly complementary to each other, implying a better result if they are combined together.

3.1.5 Prior Integrated Region Contrast

Once we obtain the region contrast map, as well as the background, depth, and orientation prior maps, we first integrate the region contrast with the background prior following the way in [33]:

$$S'_{rc}(r_i) = \sum_{j \neq i} S_{bp}(r_j) A(r_j) C(r_i, r_j), \tag{7}$$

and then simply mulitply it with the depth and orientation prior maps together to generate the saliency map:

$$S_{irc} = S'_{rc} \cdot S_{dp} \cdot S_{op}. \tag{8}$$

3.2. Global Optimization

The saliency map generated in the previous stage does not take spatial consistency into consideration. Generally speaking, neighboring superpixels are prone to have the same degree of saliency if they are of similar features. To integrate such spatial constraints, a Markov Random Field (MRF) model is employed. Moreover, based on the MRF model, we propose a guided saliency restoration scheme to

(a) Integrated Saliency (b) PageRank (c) PageRank+MRF (d) MRF (e) Ground Truth

Figure 4: A typical example for illustrating the global saliency optimization result. (a) is the saliency map produced at Stage 1. (b) is the reweighted saliency map via the PageRank algorithm. (c) is the result produced by the proposed saliency restoration scheme. (d) is the result generated by directly applying a MRF model to (a) without sampling and restoration, that is, no the sampling matrix $\mathbf{M}$ in the MRF optimization model defined in Eq.(10). (e) is the ground truth.

refine the saliency map globally. More specifically, we assume that the superpixels with high or low saliency values are, respectively, to be salient objects or the background, while the remaining may be ambiguous. We thus sample the high confident superpixels as seeds, while leaving the others as unknown. The saliency of unknown superpixles is restored through the optimization of the constructed MRF model. More details are illustrated below.

3.2.1 PageRank-based Seed Sampling

The importance of sampling representative superpixels has been realized in various works, and thus different selection schemes are proposed for optimal ranking [24, 32], saliency diffusion [20], and salient region growing [22]. In order to choose high confident superpixels, we adopt the PageRank algorithm [4, 24] to rank the salient superpixels and then sample the most top and the most bottom ones as seeds.

The PageRank algorithm reweighs the saliency of each superpixel via considering the impact from its neighborhood, by which the saliency map is updated as follows:

$$S_{pr}(r_i) = \sum_{j \in \mathcal{N}(i)} \frac{w(r_i, r_j) S_{irc}(r_j)}{\sum_{k \in \mathcal{N}(j)} w(r_k, r_j)}. \tag{9}$$

Here, the weight $w(r_i, r_j) = \langle \mathbf{n}(r_i), \mathbf{n}(r_j) \rangle$ considers the relative surface orientation of each two superpixels. This reweighting scheme implies that the saliency of a superpixel is enhanced if its neighbors are of high saliency and similar orientations, and vice versa.

3.2.2 MRF-based Saliency Restoration

A common idea shared among saliency diffusion [20], salient region growing [22], and our saliency restoration scheme is to propagate saliency from high confident elements to the unknowns. In contrast to the former two approaches, we formulate this problem within the Markov random field framework, borrowing the idea from image restoration tasks.

Our MRF model is constructed over superpixels. Let us stack the saliency of all regions as a vector $\mathbf{S} = [S_1, \cdots, S_n]$ and denote the sampled set by $\mathbf{Y} = \mathbf{MS}$. $\mathbf{M} \in \mathbb{R}^{m \times n}$ is the sampling matrix that selects m samples from the entire region set according to the seed sampling scheme. The initial $\mathbf{S} = \mathbf{S}_{pr}$, which implies $\mathbf{Y} = \mathbf{MS}_{pr}$. The task of saliency restoration aims to reconstruct the saliency map via solving

$$\mathbf{S}^\star = \arg\min_{\mathbf{S}} ||\mathbf{MS} - \mathbf{Y}||_2^2 + \alpha \sum_{i,j \in \mathcal{N}(i)} \omega_{ij} ||S_i - S_j||_2^2. \tag{10}$$

In this model, the data fidelity term enforces the consistency between the reconstructed saliency and the previous-measured value for the seeds. The pairwise smoothness term constrains the consistency between neighbors, while weighted by $\omega_{ij} = \exp(-||\mathbf{f}_i - \mathbf{f}_j||_2^2 / \sigma_{\mathbf{f}}^2)$ that incorporates color and depth information as guidance. α is a scaling factor to balance these two terms.

An example illustrated in Figure 4 shows how our global optimization performs. Compared to the MRF model without sampling matrix $\mathbf{M}$, which is shown in Figure 4d, the proposed restoration model improves the final saliency map not only on removing false salient regions but also on retaining salient regions better.

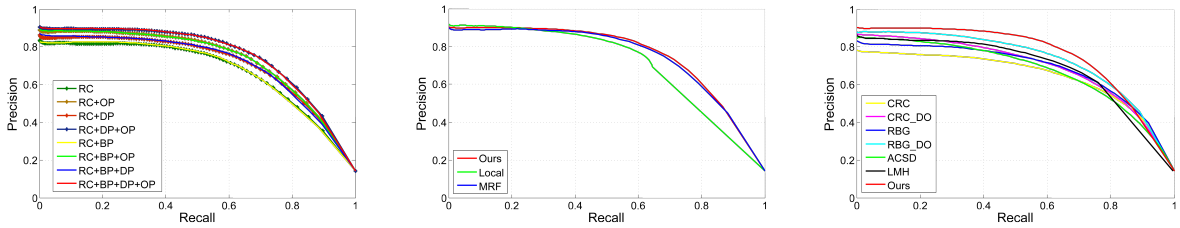
4. Experiments

4.1. Experimental Setup

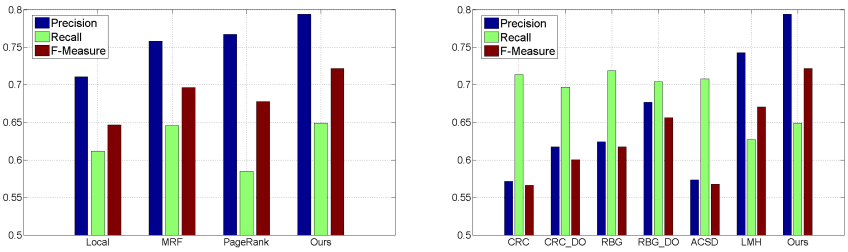
To validate our proposed approach, we have conducted experiments on two publicly available datasets: the NLPR RGBD salient object detection dataset [1] [22] and the NJU-DS400 dataset [2] [14]. The former has 1,000 natural images captured by Microsoft Kinect in diverse indoor and outdoor environments, together with labeled ground truth. The latter contains 400 stereo images, whose corresponding depth maps are generated using an optical flow method. Ground truth is also provided.

[1] https://sites.google.com/site/rgbdsaliency/dataset

[2] http://svalianju.wix.com/home#!salient-object-detection/cin4



(a) Prior integrated region contrast (b) Global optimization (c) Comparision with other methods



(d) Global optimization (e) Comparision with other methods

Figure 5: Comparative evaluation on the NLPR RGBD dataset [22]. The PR curves in (a) illustrate the performance for the combinations of different priors. (b) and (d) show the performance of the proposed global optimization scheme, compared to the one without optimization (denoted as 'Local') and the one directly applying a MRF model (denoted as 'MRF'). In (c) and (e), the performance of our methods is compared with other approaches.

During the experiments, we compare our algorithm with four state-of-the-art saliency detection methods, among which two are developed for RGB-D images and two for traditional 2D image analysis. One RGBD method [22] performs saliency detection on Low-level, Mid-level, and High-level stages and therefore is referred to as LMH here. The other RGBD method [14] is based on anisotropic center-surround difference and therefore is denoted by ACSD here. The two 2D methods are, respectively, Cheng's Region Contrast method [5] that is denoted as CRC and the approach from Robust BackGround detection [33] that is denoted as RBG. For performance evaluation, we use standard precision recall curves (PR curves) and average Precision, Recall and F-measure. Throughout all the experiments, the parameters involved in our algorithm are empirically assigned as follows: $\sigma_{\mathbf{x}} = 0.3$, $\sigma_b = 1$, $\alpha = 1$, and $\sigma_{\mathbf{f}} = 10$. The superpixels sampling rate at Stage 2 is set to be 50%. Moreover, we use Mean Shift for oversegmentation.

4.2. Experiments on NLPR RGBD Dataset

In this experiment, we first evaluate the effectiveness of each prior. By denoting region contrast as RC, background prior as BP, depth prior as DP, and orientation prior as OP, we compare the performance for all kinds of combinations. The corresponding PR curves are illustrated in Figure 5a. This comparison shows that orientation prior outperforms depth prior. The combination of these two complimentary priors improves the performance further. In addition, it turns out that the background prior does little to improvement. The main reason is that, in most cases, depth information helps to pop out the objects no matter if they are connected to the boundary or not.

Then, we evaluate the performance of our global optimization scheme. Figure 5b and Figure 5d demonstrate the PR curves and the average precision, recall and F-measure. Although no much difference is discerned from the PR curves, Figure 5d shows that the proposed optimization scheme improves the average precision and F-measure when compared to the traditional MRF model.

The comparison with other approaches is presented in Figure 5c and Figure 5e. Besides the comparison to CRC [5], RBG [33], ACSD [14], and LMH [22], for the two 2D saliency approaches, we also multiple their results with our depth and orientation prior maps, which are respectively denoted as CRC_DO and RBG_DO to investigate the effectiveness. The experiments show that both approaches are highly improved when incorporated with the two priors. Overall, our approach performs best, followed by LMH and RBG_DO. CRC and ACSD exhibit weak performance. Some typical examples are demonstrated in Figure 6.

4.3. Experiments on NJU-DS400 Dataset

To further validate our approach, we also conduct a set of comparative experiments on the NJU-DS400 dataset. In this dataset, the depth maps are produced by stereo matching and normalized to the range of [0, 1]. Therefore, it is not able to apply the orientation prior. Moreover, due to the lim-

(a) Color Image (b) Depth Map (c) CRC [5] (d) RBG [33] (e) ACSD [14] (f) LMH [22] (g) Ours (h) Ground Truth

Figure 6: Comparative results on the NLPR RGBD dataset [22].

it of stereo matching algorithms, the obtained depth values on texture-free surface are prone to be the same. Although with these problems, our approach consistently outperforms all the other methods as shown in Figure 7. ASCD, CRC_D, and RBG_D have similar performance, better than LMH, RBG, and CRC.

5. Conclusion

In this paper, we have proposed two highly complementary global priors, which are intuitive and simple for computation, to detect RGB-D salient objects. A saliency restoration scheme is also presented for refining the results globally. The proposed approach consistently outperforms other state-of-the-art algorithms on two publicly available RGB-D datasets. Moreover, the 2D background prior has also been investigated, showing that it is less effective on RGB-D data than on traditional 2D images, which is reasonable.

Acknowledgement

This work was supported by State High-Tech Development Plan (863 Program) of China (No. 2014AA09A510) and Key Program of Zhejiang Provincial Natural Science Foundation of China (No. LZ14F020003).

References

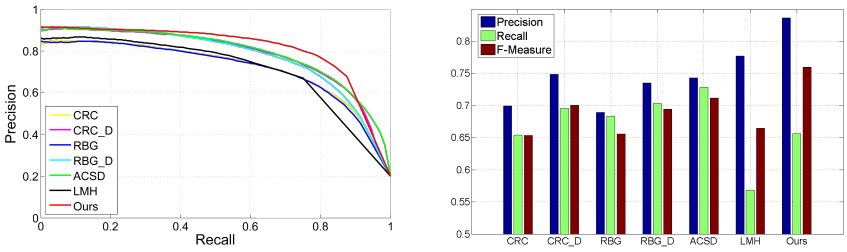
[1] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned Salient Region Detection. In *CVPR*, 2009.

[2] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Ssstrunk. SLIC Superpixels Compared to State-of-the-art Superpixel Methods. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(11):2274–2282, 2012.

[3] A. Borji and L. Itti. State-of-the-Art in Visual Attention Modeling. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(1):185–207, 2013.

[4] S. Brin and L. Page. The Anatomy of a Large-scale Hypertextual Web Search Engine. In *WWW*, pages 107–117, 1998.

[5] M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S. Hu. Salient Object Detection and Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell., Preprint*, 2015.



(a) Comparision with other methods

(b) Comparision with other methods

Figure 7: Comparative evaluation on the NJU-DS400 dataset [14]. It needs to be mentioned that the results of all other methods are produced by running the codes downloaded from the authors' website. Moreover, all the results are applied with a center-bias prior to make the final comparison.

[6] A. Ciptadi, T. Hermans, and J. M. Rehg. An In Depth View of Saliency. In *BMVC*, 2013.

[7] D. Comaniciu and P. Meer. Mean Shift: A Robust Approach Toward Feature Space Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603–619, 2002.

[8] K. Desingh, K. M. Krishna, D. Rajan, and C. V. Jawahar. Depth Really Matters: Improving Visual Salient Region Detection with Depth. In *BMVC*, 2013.

[9] J. T. Enns and R. A. Rensink. Influence of Scene-Based Properties on Visual Search. *Science*, 247(4943):721–723, 1990.

[10] D. Gao, V. Mahadevan, and N. Vasconcelos. The Discriminant Center-Surround Hypothesis for Bottom-up Saliency. In *NIPS*, 2007.

[11] S. He and R. Lau. Saliency Detection with Flash and No-flash Image Pairs. In *ECCV*, 2014.

[12] D. Hoiem, A. A. Efros, and M. Hebert. Putting Objects in Perspective. In *CVPR*, 2006.

[13] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li. Salient Object Detection: A Discriminative Regional Feature Integration Approach. In *CVPR*, 2013.

[14] R. Ju, L. Ge, W. Geng, T. Ren, and G. Wu. Depth Saliency Based on Anisotropic Center-surround Difference. In *ICIP*, 2014.

[15] C. K. L. Itti and E. Niebur. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(11):1254–1259, 1998.

[16] 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.

[17] C. H. Lee, A. Varshney, and D. W. Jacobs. Mesh Saliency. In *SIGGRAPH*, 2005.

[18] J. Li, M. D. Levine, X. An, X. Xu, and H. He. Visual Saliency Based on Scale-Space Analysis in the Frequency Domain. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(4):996–1010, 2013.

[19] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu. Saliency Detection on Light Field. In *CVPR*, 2014.

[20] R. Liu, J. Cao, Z. Lin, and S. Shan. Adaptive Partial Differential Equation Learning for Visual Saliency Detection. In *CVPR*, 2014.

[21] Y. Niu, Y. Geng, X. Li, and F. Liu. Leveraging Stereopsis for Saliency Analysis. In *CVPR*, 2012.

[22] H. Peng, B. Li, W. Xiong, W. Hu, and R. Ji. RGBD Salient Object Detection: A Benchmark and Algorithms. In *ECCV*, 2014.

[23] E. Potapova, M. Zillich, and M. Vincze. Calculation of Attention Points Using 3D Cues. In *OEAGM*, 2011.

[24] Z. Ren, Y. Hu, L.-T. Chia, and D. Rajan. Improved Saliency Detection Based on Superpixel Clustering and Saliency Propagation. In *ACM MM*, 2010.

[25] C. Rother, V. Kolmogorov, and A. Blake. GrabCut - Interactive Foreground Extraction Using Iterated Graph Cuts. *ACM TOG*, 23(3):309–314, 2004.

[26] X. Shen and Y. Wu. A Unified Approach to Salient Object Detection via Low Rank Matrix Recovery. In *CVPR*, 2012.

[27] B. W. Tatler. The Central Fixation Bias in Scene Viewing: Selecting an Optimal Viewing Position Independently of Motor Bases and Image Feature Distributions. *J. Vision*, 14(1):1–17, 2007.

[28] J. Theeuwes, P. Atchley, and A. F. Kramer. Attentional Control Within 3-D Space. *J. Exp. Psychol.*, 24(5):1476–1485, 1998.

[29] J. Wang, P. L. Callet, S. Tourancheau, V. Ricordel, and M. P. D. Silva. Study of Depth Bias of Observers in Free Viewing of Still Stereoscopic Synthetic Stimuli. *J. Eye Movement Res.*, 5(5):1–11, 2012.

[30] Y. Wei, F. Wen, W. Zhu, and J. Sun. Geodesic Saliency Using Background Priors. In *ECCV*, 2012.

[31] Q. Yan, L. Xu, J. Shi, and J. Jia. Hierarchical Saliency Detection. In *CVPR*, 2013.

[32] C. Yang, H. Lu, X. Ruan, and M. Yang. Saliency Detection via Graph-Based Manifold Ranking. In *CVPR*, 2013.

[33] W. Zhu, S. Liang, Y. Wei, and J. Sun. Saliency Optimization from Robust Background Detection. In *CVPR*, 2014.