SALIENCY ANALYSIS BASED ON DEPTH CONTRAST INCREASED

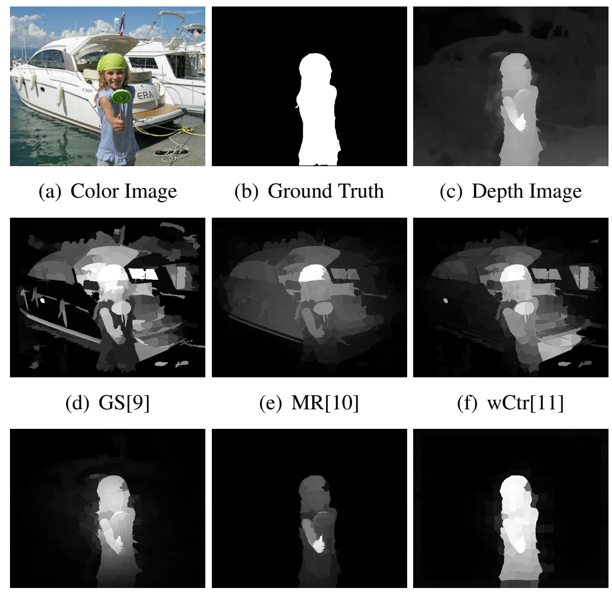
Hao Sheng[12] *Xiaoyu Liu*[1] *Shuo Zhang*[1]

[1]State Key Laboratory of Software Development Environment, School of Computer Science and Engineering, Beihang University, Beijing, P.R.China
[2]Shenzhen Key Laboratory of Data Vitalization, Research Institute in Shenzhen, Beihang University, Shenzhen, P.R. China
shenghao@buaa.edu.cn,xiaoyuliu@buaa.edu.cn,shuo.zhang@buaa.edu.cn

ABSTRACT

Humans can understand their surroundings by an additional depth cue that provides by stereopsis, which plays an important role in the human visual system. Recently, depth saliency has been attracted much attention. But depth image differs a lot from color image. Feature extraction in depth image is an important problem in depth saliency analysis. Previous studies always extract features from depth map directly. This paper proposes a method which can make the saliency analysis easier and more accurate by increasing the depth contrast between the salient object and distractors. Then, we extended a recent saliency analysis approach to evaluate the saliency of the difference map. Finally, after the optimization by depth information and color information the final saliency map can be obtained. Our experiments on public dataset show that our method significantly outperforms state-of-the-art.

***Index Terms*—** saliency detection, depth information, increased depth contrast



(a) Color Image (b) Ground Truth (c) Depth Image
(d) GS[9] (e) MR[10] (f) wCtr[11]
(g) ACSD[12] (h) SS[13] (i) Ours

Fig. 1. Saliency maps. When an object and distractors share some common visual attributes, the intrinsic cues are often insufficient to distinguish them. This example shows that saliency analysis based on depth information can be a useful complement to existing visual saliency.

1. INTRODUCTION

In recent years, image saliency[1] has attracted much attention. As studied in cognitive psychology[2] and neuroscience[3], visual attention for human including visual information gathering and filtering that makes something saliency in an image catch a viewers attention immediately. In computer vision problems, saliency detection[4] is an important preprocessing step to reduce computational complexity in the early stage of many higher visual analysis applications, such as object segmentation[5], image classification[6, 7], image compression[8] and so on.

In the past decades, a lot of approaches have been proposed for detecting salient objects in images[4, 14, 15, 16, 17]. How to distinguish salient objects from background is the key step of saliency detecting. In existing works, there are mainly two kinds of approaches to solve this problem. A wide variety of computational methods have been developed to estimate saliency from intrinsic cues[18, 19]. Most of these methods compute saliency based on feature only from the input image itself such as color, shape, texture, etc. However, in many cases the intrinsic cues are often insufficient to distinguish salient object from distractors. So another kind of approaches proposes to calculate saliency via extrinsic cues such as depth map[20, 12, 13], user annotations, or image statistical information. Therefore, methods based on extrinsic cues often using the intrinsic cues at the same time. As is shown in Figure 1, the three methods in the second raw are based on only intrinsic cues and the three methods in the third raw are based on stereo and depth information. It is hard to distinguish objects and distractors only from intrinsic cues, because they share some common visual attributes. At the same time, depth information supplies a powerful cue for saliency detection.

Human beings live in a 3D world and human visual system operates in real 3D environments. Humans can understand their surroundings by an additional depth cue that provides by stereopsis, which plays an important role in the human visual system. Similar to color image, feature extraction is also an important problem in depth saliency analysis. Studies using depth information for saliency detection always extract features from depth map directly[20, 12]. But as depth image differs a lot from color image, is there a better way to extract features from depth map?

To solve this problem, this paper proposes a method which can make the saliency analysis easier and more accurate by increase the depth contrast between the salient object and background. This method is based on the following considerations: 1. The area of salient regions in the whole image will not be large. 2. When taking a photo, people tends to place an important object at a different depth level than the others. We get a surface which can considered as most of the original depth map points on it by fitting the input depth map. Then by using the original depth map subtract this surface, a difference map can be got, which increasing contrast between salient object and background. Then, we extended a recent saliency analysis approach which based on manifold ranking to evaluate the saliency of the difference map. Finally, after the optimization by depth information and color information the final saliency map can be obtained.

2. METHOD

2.1. Depth Contrast Increased

This paper presents a novel method to make depth saliency easy to evaluate. Firstly, we show the influence of different depth contrast. As Fig. 2 shows, due to the restriction of environment factors, photographer always takes photo on view point 1. On such a view point, the obtained RGB image is shown as Fig. 2(b) and the depth image is shown as Fig. 2(d). In this situation, the depth distribution is very wide and the contrast between each part of image is not clear. And RGB and depth image like Fig. 2(c) and Fig. 2(e) can be obtained when the photographer takes photo on view point 2. Comparing the four pictures, Fig. 2(b) and Fig. 2(c) contains almost same information. But there are large differences between these two depth images. To Fig. 2(d), a very poor saliency map can be got by using the contrast-based saliency method. But to Fig. 2(e), a very perfect saliency map can be obtained by using the same method. It is because that compared with the global depth, depth of object in Fig. 2(e) is more outstanding than it in Fig. 2(d). To photos which are shooting on view point 2, the saliency is very easy to evaluate. But unfortunately, most of the photos are taken on view point 1.

According to the above phenomenon, if we can change

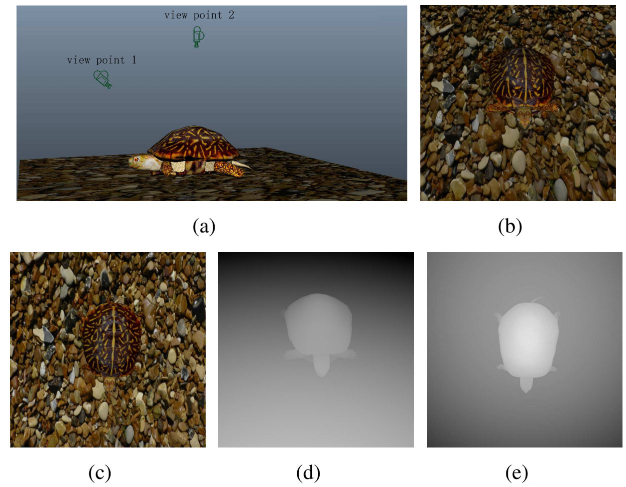


Fig. 2. This example shows the difference between RGB and depth images of different view points.

from view point 2 to view point 1, then the evaluation of saliency will be very easy. But the only information can be got is the image itself, so there is no way to carry out the view point change in fact. Based on the domain knowledge in photography that important objects always placed at a different depth level than the others things in the scene and salient regions always take small areas. This paper presents a simple method to increase depth contrast and get a good result like view point change. A surface can be achieved by using this surface to fit the input depth map. In order to facilitate understanding, this surface can considered as a surface which has most of the original depth map points on it. Then using the original depth map subtract this surface, a difference map which contrast between salient object and backgrounds has be increased can be got. As shown in Fig. 3, comparing with depth image, object regions in the difference image has a clear contrast with other regions. Moreover, in some ideal cases, the difference map can pop-out the object completely.

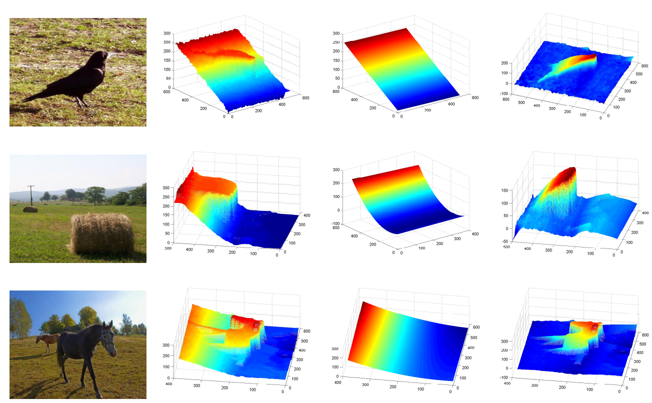


Fig. 3. From left to right: Color Image, Depth Image, Fitting Image and Difference Image.

In practice, we increase the contrast between salient object and backgrounds by following steps. Firstly, by using pixel value of the depth map as dependent variable and image coordinate value as independent variable, linear regression based on least square algorithm can be used to fit a new plane. This process can be understood as founding a plane which contains most depth map points on it. For each pixel, a new value can be obtained by using the original depth value minis the fitting value. Secondly, our real world is three-dimensional, so the background in many scenes is not simply a plane. In fact, in many saliency dataset, most of the scenes have a complex background. In this situation, objects which popping out from the background tend to be salient is still tenable. So by using our method, the goal of reduces the difficulty of saliency calculation still can be achieved. But it is impossible to adapt a variety of situations with a plane background template. In order to maximize the contrast between salient object and backgrounds, quadric surface was used to solve this problem. So the process of first step can be repeated, but use x^2, y^2, xy, x, and y as independent variable instead. As a result, another difference map can be got. Finally, it is difficult to combine multiple saliency cues, as many previous works, we simply summation these two difference map as the final difference map.

2.2. Depth Saliency Analysis and Optimization

This paper extend a recent color based saliency detection method from Yang[10] for depth saliency analysis. This is a bottom-up method which measure the saliency of a pixel or region based on its low-level cues. And it is based on the consideration that center regions are more salient than peripheral, which has been demonstrated by previous work[21]. In this method image pixels or regions are ranked based on the similarity to background and foreground queries via manifold ranking. Manifold ranking is a data classification method, which assigns ranks to elements in a data set. It can reveal their likelihood being in a certain group with respect to the intrinsic manifold structure. A good survey can be found in [22]. When we use this method, the weight of the edge is computed based on the distance in the depth space and is defined as the following equation :

$$w_{ij} = \exp\{-\frac{\|d_i - d_j\|^2}{\sigma^2}\} \tag{1}$$

In this equation d_i and d_j are the mean value of the two superpixels, and σ is controlling constant of the weight strength. Then, we can use this weight matrix to calculate a saliency map.

There are a lot of optimization methods, but rarely have methods considered depth feature and RGB feature simultaneously. This paper use a saliency optimization method to integrates depth cues and RGB cues. In order to achieve this goal, the cost function blow will be minimized.

$$f(s_i) = \sum_{i=1}^{N} fb_i(s_i - 1)^2 + \sum_{i=1}^{N} pb_i s_i^2 + \sum_{i,j=1}^{N} w_{ij}(s_i - s_j)^2 \tag{2}$$

This function control the optimization from three aspects. The first term imply nodes which have large foreground probability tend to have a large saliency value. The foreground probability fb in the function is result of previous stage. The second term encourage nodes with large background probability have a less saliency value. In this term, the background probability is initialized by 1 minus foreground probability, then, optimized by the depth prior which far regions appear large background probability. The third term use color information and encourage adjacent regions with similar color have similar saliency, which can smooth the saliency map by decrease the noise in both background and foreground. The weight w_{ij} is defined as Gaussian weighted function like equation 1, but the distance between two nodes is calculated in CIE-Lab color space. Then we can minimize the cost function though least-square which takes very little time and obtain a optimized saliency map.

3. EXPERIMENT AND ANALYSIS

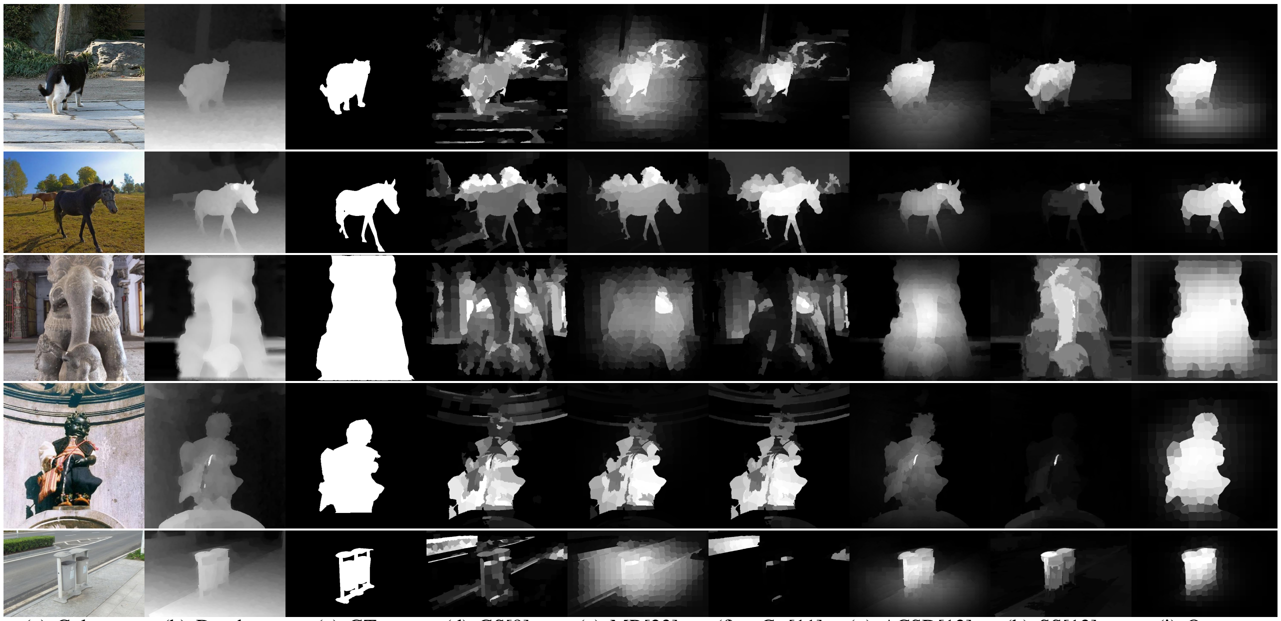
3.1. Dataset

There are many public benchmark datasets available for RGB saliency analysis, but few publicly dataset with depth information. To the best of our knowledge, we found two datasets: a stereo saliency analysis benchmark dataset that contains 1000 stereoscopic image built by [13] and NJU-2000 dataset which has 2000 images with depth information provide by [12]. In these two datasets, most color images are stereo images collected from Internet. So, depth or disparity information can be got through stereo matching and the ground truth can be labeled following the procedure in [13]. Here, we use NJU-2000 dataset to evaluate our algorithm.

3.2. Result Analysis

For performance evaluation, we use the widely used precision-recall curves (PR curves). Specifically, we obtain a curve for every image by comparing the ground truth and a series of binary masks which can be got from the saliency map by using increasing threshold from 0 to 255. Then, the curves are averaged on dataset. We compare with the most recent five state-of-the-art saliency detection methods, including MR[23], wCtr[11], GS[9], SS[13] and ACSD[12]. Among these, the first three methods are based color images; SS[13] uses the disparity information and stereoscopic rules; ACSD[12] works on depth images.

We show some example saliency maps generated by previous methods in Fig. 4. It can be see that, methods based



(a) Color (b) Depth (c) GT (d) GS[9] (e) MR[23] (f) wCtr[11] (g) ACSD[12] (h) SS[13] (i) Ours

Fig. 4. Saliency map of different methods. The first column shows the input color images. The second and third column shows the depth images and ground truth salient object masks. The next three columns shows the results of color image based methods. The last three columns are the saliency results of depth based methods

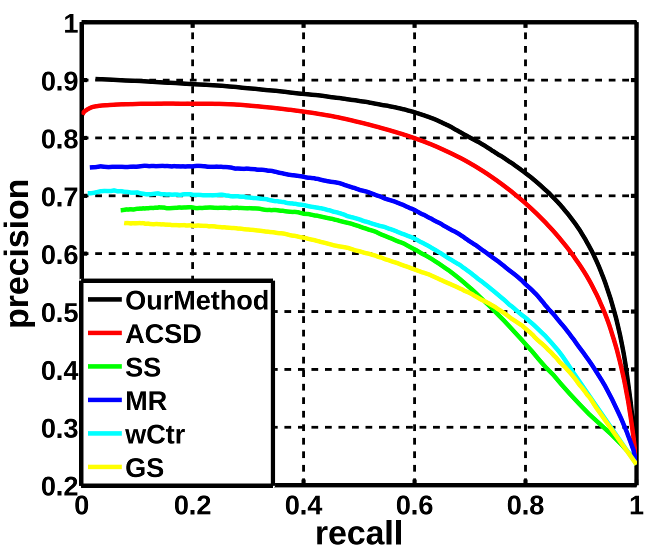


Fig. 5. Precision-recall curves of several methods.

on depth information perform better than color information based methods. And, the saliency maps generated by our method have clearer details and finer boundaries. Moreover, by using our increased depth contrast method, we can handle some case (like the fourth image in Fig. 4) that other depth based methods can't. The precision-recall curves are given in Fig. 5. Again, this paper proposed method outperforms all of the other approaches.

4. CONCLUSION

This paper explored a new way to use depth information. We present a novel and simple method to get a contrast increased map by using the original depth map subtracts the fitted surface. Then, we extended a recent saliency analysis approach to evaluate the saliency of the generated difference map. Finally, both depth and color information are used to optimize the saliency map. Extensive experimental results demonstrate that the proposed method significantly outperforms 5 state-of-the-art saliency detection algorithms and our depth contrast increase method is a useful complement to existing saliency analysis.

5. ACKNOWLEDGMENT

This study was partially supported by the National Natural Science Foundation of China (No. 61472019), the National High Technology Research and Development Program of China (No. 2013AA01A603) and the National Science & Technology Pillar Program (No.2015BAF14B01). Supported by the Programme of Introducing Talents of Discipline to Universities and the Open Fund of the State Key Laboratory of Software Development Environment under grant #SKLSDE-2015KF-01.

6. 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 & Machine Intelligence*, pages 1254–1259, 1998.

[2] A. M. Treisman and G Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12, 1980.

[3] L. Itti and C. Koch. Computational modelling of visual attention. nature reviews. In *Neuroscience*, pages 367–379, 2001.

[4] Ming Cheng, Niloy J Mitra, Xumin Huang, Philip HS Torr, and Song Hu. Global contrast based salient region detection. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, pages 569–582, 2015.

[5] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. Grabcut – interactive foreground extraction using iterated graph cuts. *Acm Transactions on Graphics*, 23(3):309–314, 2004.

[6] Gaurav Sharma, Frdric Jurie, and Cordelia Schmid. Discriminative spatial saliency for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3506–3513, 2012.

[7] Lai Kuan Wong and Kok Lim Low. Saliency-enhanced image aesthetics class prediction. In *Proceedings / ICIP ... International Conference on Image Processing*, pages 997–1000, 2009.

[8] Itti Laurent. Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Transactions on Image Processing*, 13(10):1304–1318, 2004.

[9] Yichen Wei, Fang Wen, Wangjiang Zhu, and Jian Sun. Geodesic saliency using background priors. In *European Conference on Computer Vision*, pages 29–42. Springer, 2012.

[10] Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang. Saliency detection via graph-based manifold ranking. In *Computer Vision and Pattern Recognition*, pages 3166–3173. IEEE, 2013.

[11] Wangjiang Zhu, Shuang Liang, Yichen Wei, and Jian Sun. Saliency optimization from robust background detection. In *Computer Vision and Pattern Recognition*, pages 2814–2821. IEEE, 2014.

[12] Ran Ju, Ling Ge, Wenjing Geng, Tongwei Ren, and Gangshan Wu. Depth saliency based on anisotropic center-surround difference. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 1115–1119, 2014.

[13] Yujie Geng. Leveraging stereopsis for saliency analysis. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 454–461, 2012.

[14] Yao Qin, Huchuan Lu, Yiqun Xu, and He Wang. Saliency detection via cellular automata. In *Computer Vision and Pattern Recognition*, pages 110–119, 2015.

[15] Na Tong, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang. Salient object detection via bootstrap learning. In *Computer Vision and Pattern Recognition*, pages 1884–1892, 2015.

[16] Changyang Li, Yuchen Yuan, Weidong Cai, Yong Xia, and David Dagan Feng. Robust saliency detection via regularized random walks ranking. *Computer Vision and Pattern Recognition*, pages 2710–2717, 2015.

[17] Radhakrishna Achanta, Francisco Estrada, Patricia Wils, and Sabine Süsstrunk. Salient region detection and segmentation. In *Computer Vision Systems*, pages 66–75. Springer, 2008.

[18] Federico Perazzi, Philipp Krähenbühl, Yael Pritch, and Alexander Hornung. Saliency filters: Contrast based filtering for salient region detection. In *Computer Vision and Pattern Recognition*, pages 733–740. IEEE, 2012.

[19] Yulin Xie, Huchuan Lu, and Ming-Hsuan Yang. Bayesian saliency via low and mid level cues. *IEEE Transactions on Image Processing*.

[20] Congyan Lang, Tam V. Nguyen, Harish Katti, Karthik Yadati, Mohan Kankanhalli, and Shuicheng Yan. Depth matters: Influence of depth cues on visual saliency. *Lecture Notes in Computer Science*, (2):101–115, 2012.

[21] Ran Margolin, Avishay Tal, and Lihi Zelnik-Manor. What makes a patch distinct? In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1139–1146, 2013.

[22] Dengyong Zhou, Jason Weston, Arthur Gretton, Olivier Bousquet, and Bernhard Sch?lkopf. Ranking on data manifolds. In *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 16*, pages 169–176, 2003.

[23] Bowen Jiang, Lihe Zhang, Huchuan Lu, Chuan Yang, and Ming-Hsuan Yang. Saliency detection via absorbing markov chain. In *International Conference on Computer Vision*, pages 1665–1672. IEEE, 2013.