RELATIVE LOCATION FOR LIGHT FIELD SALIENCY DETECTION

Hao Sheng[1,2] *Shuo Zhang*[1] *Xiaoyu Liu*[1] *Zhang Xiong*[1,2]

[1] State Key Laboratory of Software Development Environment, School of Computer Science and Engineering, Beihang University

[2] Shenzhen Key Laboratory of Data Vitalization, Research Institute in Shenzhen, Beihang University, Shenzhen, P.R. China

ABSTRACT

Light field images, which capture multiple images from different angles of a scene, have been proved that can detect salience regions more effectively. Instead of estimating depth labels from the light field image first, we proposed to extract a relative location from the raw image for saliency detection more simply. The relative location is calculated by comparing the raw light field image captured by a plenoptic camera and the central view of the scene, which can distinguish whether the object is located before the focus plane of the main lens or not. The relative location is then integrated to a modified saliency detection framework to obtain the salience regions. Experimental results demonstrate that the proposed relative location can help to improve the accuracy of results, and the modified framework outperforms the state-of-the-art methods for light field images saliency detection.

***Index Terms*—** Light field, Saliency detection, Relative location, Raw image, Plenoptic camera

1. INTRODUCTION

Since saliency detection technology has been well developed these years, extracting salient objects from different kinds of images has also attracted much attention. Except the color, shape, and texture information acquired from traditional cameras, the structure information calculated from Kinect or binocular camera has been proved that can better improve the saliency detection results [1, 2, 3].

However, it used to difficult to capture structure information for saliency detection untill handheld light field cameras, *e.g.* Lytro [4] and Raytrix [5], appeared. Different from traditional stereoscopic images, the light field captures images of the scene from different continues angels. This character makes the camera able to extract structure features easily.

Saliency detection from light field images has been studied in [6] recently, which effectively prove that the light field images is able to detect difficult salient object. Similar to saliency detection using binocular images, the depth map is need to before the saliency detection. Although the disparity estimation from multiple images has been a historical problem for a long time, and various advanced technologies has been proposed, it is still a demanding problem for saliency detection.

Different from [6] and other saliency analysis on stereoscopic images [7, 1], we propose to use the inherent structure information in light field images for better saliency detection. Instead of figuring out the accurate depth value of every pixels of the images, we try to utilize the relative location as a feature to distinguish different locations of objects. In other words, we extract specific features directly from the raw images of plenoptic light field cameras. Moreover, we modified the traditional saliency detection work [8] to utilize the extracted features from light field images and achieve a comparable results with the the state-of-the-art methods.

2. RELATED WORK

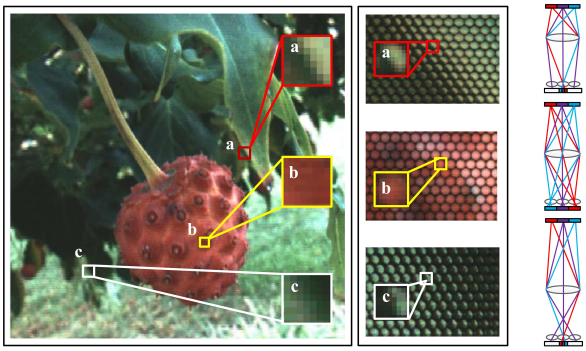
The saliency detection from light field involves how to extract depth cues from the images and how to integrate this cues with color, textures and other features for saliency analysis. Some prior work has been done for these problems.

2.1. Saliency detection using depth cues

The prior work about how to leverage depth to facilitate the saliency analysis has been discussed in [1, 2, 3]. The dataset includes the RGBD images from Microsoft Kinect and binocular camera. Depth images of the binocular images are calculated in advance using the common depth estimation methods [9, 10]. Their works focused on how to integrate the depth cues with appearance cues for saliency estimation, and proposed hypothesis reasonably and accurately.

2.2. Depth estimation from light field

Recently, some depth estimation methods has been developed specially for light field images [11, 12, 13]. The continuous sampling in angular domain is different from the traditional depth estimation for multiple view images. Because the different views of the scene can be obtained, the disparity can be estimated using stereo match [12]. On the other hand, averaging the pixels in different arrangement is able to construct the images that focuses in different depth [14]. The disparity is then acquired through measure the focusness of each images.



(b) Scene image closeup

Fig. 1. The Scene image and the angular sampling image of occluded points at the right depth. The angular sampling image always show the similar structure of the scene when the point is occluded by another object.

However, both methods rely on the assumption of the discrete depth label, which is a time-consuming progress. Moreover, due to the heavy noises and spatial aliasing [15] in plenoptic light field images, these methods have difficulty in estimating depth in light field images like Lytro.

2.3. Saliency detection for light field

The saliency detection for light field images is first proposed in [6]. They first calculate the focus stack using the refocus theory [14], and then estimate the in-focus regions in every images. The depth map is then obtained, and combined with the objectness to estimate the foreground likelihood and background likelihood. Their work proves that the additional information in light field images can contribute to saliency detection. However, the refocus process and the in-focus region estimation needs to be calculated many times to acquire the relative depth map, which is a time-consuming process.

In contrast, our approach does not try to calculate the complete depth map. Instead, we develop a simple method to utilize the structure difference specific to our light field dataset for saliency detection. The relative locations with respect to the focused plane of the main lens is calculated and integrated to acquire the salience map.

3. RELATIVE LOCATION EXTRACTION

In this paper, we use $L(x, y, u, v)$ to parametrize the 4D light field, where (u, v) is the coordinate of the main lens plane and (x, y) is the coordinate of the image in different views. The raw image of Lytro is shown in Fig.1, the circular region is the image under each micro-lens. If we pick up the corresponding pixels in every micro-lens, we can acquire one view image of the scene. We do not need to calculate the matching cost at different depth level [] or calculate the focusness in every images, which are time-consuming. By contrast, we aim at finding the relative location relationship which is sufficient to distinguish the background and foreground in saliency detection.

3.1. Background and foreground filters

Due to the construction of the camera, the image under each micro-lens is closely related to the position in the scene. As shown in Fig.1, if the scene is behind the focusing plane, the micro-lens image is . On the contrary, if the scene is before the focusing plane, the image is . if the object is located at the focusing plane, the picture is just .

Based on the observations, we construct a specific feature to present whether the point is before, behind or just on the focusing plane of the main lens. We build two filters, foreground filter w_{ff} and background filter w_{bf}, to evaluate the possibilities of the points' position.

We first define a general linear filter as the popular bilateral [16] or guided filter [17], which treats a view image I_v as a guidance image, the raw light field image I_r as an input image. The output image of the filter $W_{i,j}$ is expressed as a weighted average of each micro-lens image:

$$I(q_i) = \sum_j W_{ij}(I_v)(I_r(p_j) - I_r(p_i)), \tag{1}$$

where p_i is the center pixel of each micro-lens image, and q_i is the corresponding output which has the same size as the view image.

The foreground filter W^f is constructed according to the view image:

$$W^f_{ij} = exp(-\frac{|x_i - x_j|^2}{2\delta^2}), \tag{2}$$

and background filter W^b is set as the transpose of the W^f. The window size is set according to the minimum and maximum depth respectively. In this paper, we set the window size equal to the size of the micro-lens to fit for the filtering in Equ. 1. The experiments in the realistic scene prove that it is sufficient for the depth range.

3.2. Relative Location

The two filter is then applied to the raw images and obtain the filtered result image I^f and I^b. If the point is behind the focusing plane, I^f is larger than I^b. On the contrary, if the scene is before the focusing plane, I^f is larger than I^b. If the object is located at the focusing plane, I^f has the approximate value as I^b.

In order to remove noises and propagate the credible information, we filter the I^f and I^b using guided filter [17], and then the relative location is defined as:

$$L = \frac{I^f - I^b}{I^f + I^b}, \tag{3}$$

where $I_d \gg 1$ indicates the possibilities of the depth position is behind the focusing plane, $I_d \ll 1$ before the focusing plane on the contrary, and $I_d \approx 1$ on the focusing plane.

4. SALIENCY DETECTION

In this section, we detect the salince part of the scene using the extracted relative location combined with the color information. We show how to integrate all the information and obtain the final salience map. The entire salince detection frame is based on the work of Zhu *et al.* [8], and we add the relative location cues calculated in the last section to better detect the salince part. First, We segment the reference image into a set of superpixels using mean-shift algorithm [18]. The relative location cues is then computed as the average value of all pixels within a region $d(p)$.

4.1. Background Selection

In [8], the robust background measure assumes that the object regions are much less connected to image boundaries than background ones. However, if the background is complex, as shown in Fig.3, we cannot effectively determine whether they are connect to image boundaries based on the color information. Hence, the relative location cues are added.

An undirected weighted graph is first constructed. All adjacent superpixels (p, q) are connected and their weight $d(p, q)$ is assigned as the Euclidean distance between their average colors and relative location:

Then the boundary connectivity is defined:

$$BndCon(p) = \frac{Len_{bnd}(p)}{\sqrt{Area(p)}}, \tag{4}$$

where the definition of $Len_{bnd}(p)$ is the length along the boundary and $Area(p)$ is a soft area of the region that p belongs to. The detail definitions are similar with [8] except that the weight $d(p, q)$. The $d(p, q)$ not only consider the color information, but also fuse the location information. This setting can effectively connect the background to the image boundaries whether the color of the background is complex, or the depth of the background is changing. The former one is also a common problem in RGBD image salincy deteion.

Then the background is selected as:

$$\omega_i^{bg} = 1 - exp(-\frac{BndCon^2(p_i)}{2\sigma^2_{bndCon}}), \tag{5}$$

4.2. Contrast Selection

Most RGBD salince detection work defined that the object which is closer to the camera is more likely to be salince. This assumption is partly correct except two common scenes. First, the overall location of the object is close to the camera, but it is connected closely to the image boundaries. Secound,

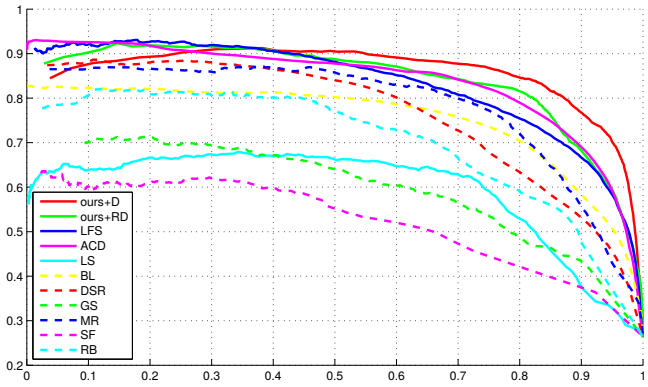


Fig. 3. Comparison of our saliency maps with state-of-the-art methods.

the depth of the object is changing sharply in the image, *e.g.* the ground or the flat desktop. In order to remove the effects of the two scenes, we calculate the contrast as:

$$\omega_i^{fg} = \sum_{i=1}^{N} d(p, q)\omega_{spa}(p, p_i)\omega_i^{bg} Area(p), \tag{6}$$

where $\omega_{spa}(p, p_i)$ is define as in [8]. The $Area(p)$ is calculated according to the relative location cues. On one hand, if the overall location of the object is close to the camera, and it is connected closely to the image boundaries, the ω_i^{bg} will be large. On the other hand, if the depth of the object is changing sharply, the soft area $Area(p)$ is averaged to be relative lower than the other objects.

Finally, the salience map is optimized by minimize cost function:

$$\sum_{i=1}^{N} \omega_i^{bg} s_i^2 + \sum_{i=1}^{N} \omega_i^{fg} (s_i - 1)^2 + \sum_{i,j} \omega_{i,j}(s_i - s_j)^2 \tag{7}$$

as in [8], the optimal salince map is computed by least-square.

5. EXPERIMENT

In this section, a dataset of 100 light field images [6] is used to evaluate the proposed method. We compare our method with state-of-the-art salience detection methods desgined for light field image (LF [6]), RGBD images (ACD [7], LS [1]) and traditional RGB images (RB [8], BL [19], DSR [20], GS [21], MR [22], SF [23]). We evaluate our experimental results using both relative location cues (RD) and depth maps (D) to show the effectiveness of the relative location cues and the proposed saliency detection method. The depth maps used for RGBD salince detection are calculated using the depth from focusness method [24], which are also released in the dataset [6].

The visual examples are shown in Fig. 2. We can observe that the relative location cues are able to distinguish the

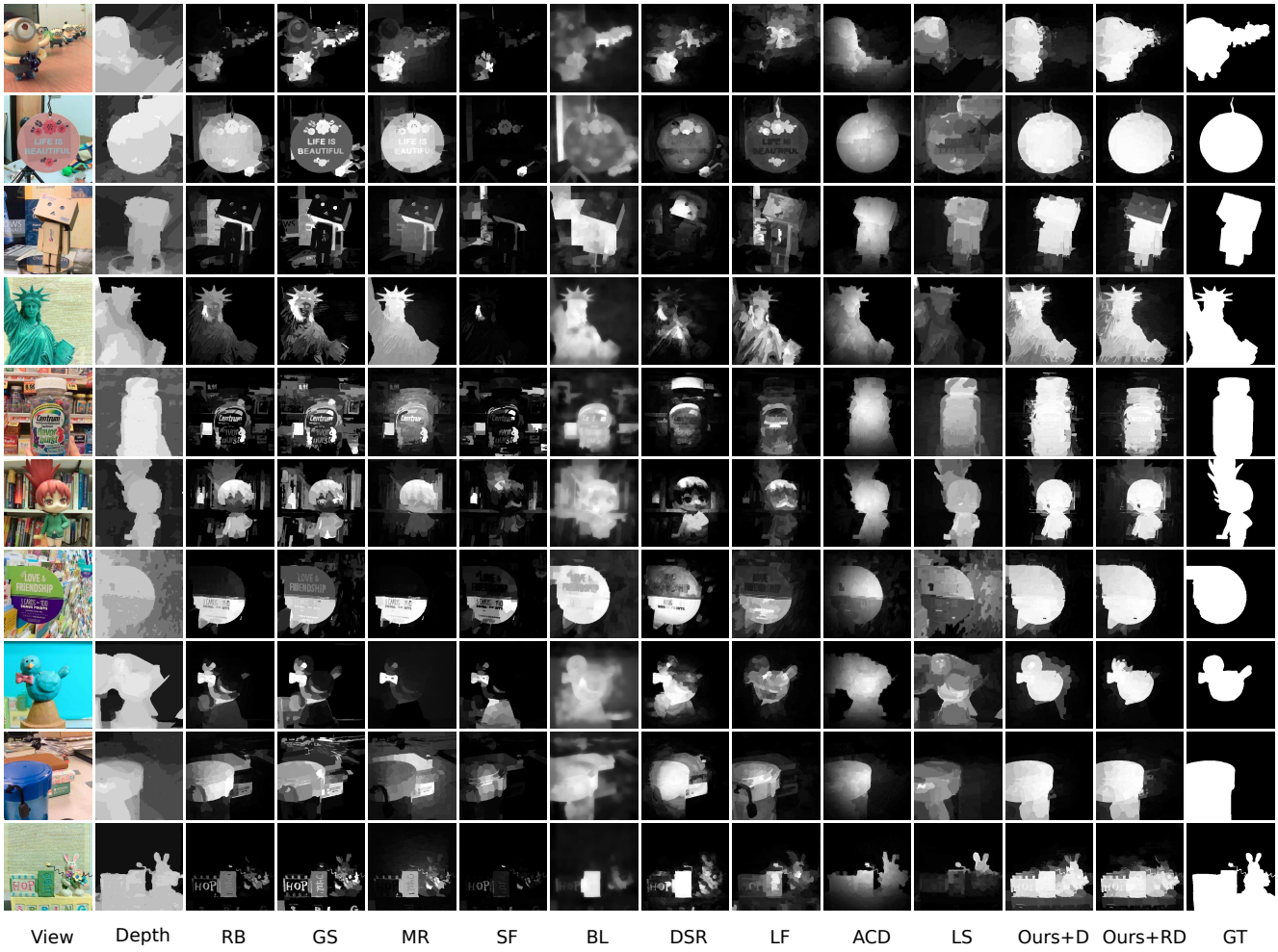


Fig. 2. Precision recall curve comparision with state-of-the-art methods.

outstanding objects clearly and highlight the salience parts. Compared with LF [6], the saliency parts are more outstanding because of the simple relative location. We can also verify the effectiveness of the modified saliency detection method by using the RGBD images, as comparing with ACD [7] LS [1].

We also calculate the precision-recall curve (PRC) in Fig.3 to show the similarity between the detected salience map and the ground truth. We binarize the saliency map at each possible threshold with in [0, 255]. As we can see in the figure, the proposed method using RGBD images achieves a higher precision and recall rate compared with using the relative location cue. The reason is that the depth map is more precise than the relative location. However, it also has a more complex calculation. As a result, we can chooes different cues based on different requirements for light field saliency detection.

6. CONCLUSION

Taking into account the special structure of the light field images, we propose a novel relative location cues to extract the salience parts of an image. The relative location is calculated on the raw images, which is simple and effective. Based on the locations with respect to the focused plane, we can extract the salience regions using a modified salience detection method. The information is then integrated to highlight the objects which are closer with the camera. Compared with the state-of-the-art methods, the proposed method is able to detect salience more precisely as well as simply. Moreover, the proposed saliency detection framework is also proved to be adapted to the RGBD images.

7. ACKNOWLEDGMENT

This study was partially supported by the National Natural Science Foundation of China (No.61272350), 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-2015ZX-21.

8. REFERENCES

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

[2] Houwen Peng, Bing Li, Weihua Xiong, Weiming Hu, and Rongrong Ji, "Rgbd salient object detection: a benchmark and algorithms," in *Computer Vision–ECCV 2014*, pp. 92–109. Springer, 2014.

[3] Yuming Fang, Junle Wang, Manish Narwaria, Patrick Le Callet, and Weisi Lin, "Saliency detection for stereoscopic images," in *Visual Communications and Image Processing (VCIP), 2013*. IEEE, 2013, pp. 1–6.

[4] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR*, vol. 2, no. 11, 2005.

[5] Christian Perwa and Lennart Wietzke, "The next generation of photography," http://www.raytrix.de.

[6] Nianyi Li, Jinwei Ye, Yu Ji, Haibin Ling, and Jingyi Yu, "Saliency detection on light field," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 2806–2813.

[7] 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*. IEEE, 2014, pp. 1115–1119.

[8] Wangjiang Zhu, Shuang Liang, Yichen Wei, and Jian Sun, "Saliency optimization from robust background detection," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 2814–2821.

[9] Ce Liu, Jenny Yuen, and Antonio Torralba, "Sift flow: Dense correspondence across scenes and its applications," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 5, pp. 978–994, 2011.

[10] Daniel Scharstein and Richard Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International journal of computer vision*, vol. 47, no. 1-3, pp. 7–42, 2002.

[11] Sven Wanner and Bastian Goldluecke, "Globally consistent depth labeling of 4d light fields," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012, pp. 41–48.

[12] Can Chen, Haiting Lin, Zhan Yu, Sing Bing Kang, and Jingyi Yu, "Light field stereo matching using bilateral statistics of surface cameras," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[13] Zhan Yu, Xinqing Guo, Haibing Ling, Andrew Lumsdaine, and Jingyi Yu, "Line assisted light field triangulation and stereo matching," in *IEEE International Conference on Computer Vision (ICCV)*, 2013, pp. 2792–2799.

[14] Marc Levoy, "Light fields and computational imaging," *IEEE Computer*, vol. 39, no. 8, pp. 46–55, 2006.

[15] Tom E Bishop and Paolo Favaro, "Plenoptic depth estimation from multiple aliased views," in *IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops)*. IEEE, 2009, pp. 1622–1629.

[16] Carlo Tomasi and Roberto Manduchi, "Bilateral filtering for gray and color images," in *IEEE International Conference on Computer Vision (ICCV)*. IEEE, 1998, pp. 839–846.

[17] Kaiming He, Jian Sun, and Xiaoou Tang, "Guided image filtering," in *European Conference on Computer Vision (ECCV)*, pp. 1–14. Springer, 2010.

[18] Dorin Comaniciu and Peter Meer, "Mean shift: A robust approach toward feature space analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 5, pp. 603–619, 2002.

[19] Na Tong, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang, "Salient object detection via bootstrap learning," in *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, 2015, pp. 1884–1892.

[20] Xiaohui Li, Huchuan Lu, Lihe Zhang, Xiang Ruan, and Ming-Hsuan Yang, "Saliency detection via dense and sparse reconstruction," in *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013, pp. 2976–2983.

[21] Yichen Wei, Fang Wen, Wangjiang Zhu, and Jian Sun, "Geodesic saliency using background priors," in *Computer Vision–ECCV 2012*, pp. 29–42. Springer, 2012.

[22] 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 (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 3166–3173.

[23] Federico Perazzi, Philipp Krähenbühl, Yael Pritch, and Alexander Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 733–740.

[24] Fatih Porikli Feng Li (Mitsubishi Electric Res. Labs), "Harmonic variance: A novel measure for in-focus segmentation," in *Proceedings of the British Machine Vision Conference*. 2013, pp. 33.1–33.11, BMVA Press.