

RELATIVE LOCATION FOR LIGHT FIELD SALIENCY DETECTION

Hao Sheng^{1,2} Shuo Zhang¹ Xiaoyu Liu¹ 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 salient regions more effectively. Instead of estimating depth labels from light field images, we proposed to extract relative locations, which can distinguish whether the object is located before the focus plane of the main lens or not, for saliency detection. The relative locations are calculated by comparing raw light field images captured by plenoptic cameras and central views of scenes. The relative locations are then integrated to a modified saliency detection framework to obtain the salient regions. Experimental results demonstrate that the proposed relative locations can help to improve the accuracy of results, which is also efficient. Moreover, 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 be difficult to capture structure information of a scene for saliency detection until handheld light field cameras, *e.g.* Lytro [4] and Raytrix [5], appeared. Due to the optic construction, this kind of camera is able to capture the scene from different continuous angles in one shot. Therefore, light field images contain structure information which can be extracted easily for further processes.

Saliency detection from light field images has been studied by Li *et.al* recently, which effectively proved that light field images are able to detect salient objects in complex environments [6]. Similar to saliency detection using binocular images, depth maps are needed during the saliency detection calculation in their work. Although the depth estimation from light field images has been a historical problem for a long

time, and various advanced technologies have been proposed, it is still a demanding problem for saliency detection.

Different from Li *et.al* [6] and other saliency analysis on stereoscopic images [7, 1], we proposed to use the inherent structure information in light field raw images for saliency detection. Instead of figuring out the accurate depth value of each pixel in the image, we develop a new feature, the relative location, to distinguish objects locations related to the focus plane. The proposed feature can be calculated directly from the raw images of plenoptic light field cameras. Moreover, we modified the traditional saliency detection work [8] to utilize the extracted feature and achieve comparable results compared with the state-of-the-arts methods.

2. RELATED WORK

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

2.1. Saliency detection using depth cues

The work about how to leverage depth to facilitate the saliency analysis has been discussed in [1, 2, 3]. The dataset includes RGBD images from Microsoft Kinect and binocular cameras. Depth information of the binocular images is calculated in advance using the common depth estimation methods [9]. Their work focused on how to integrate the depth cues with appearance cues for saliency estimation and proposed the depth hypothesis reasonably and accurately. However, the process of depth cues is still complex and time-consuming for the saliency detection.

2.2. Depth estimation from light field

Recently, some depth estimation methods have been developed specially for light field images [10, 11]. Some methods are designed based on stereo matching [11, 12] which use the extracted multi-views from light field. Besides, depth maps can also be calculated by measuring the focusness of each refocused image [13] from light field images. However, both methods rely on the assumption of discrete depth labels,

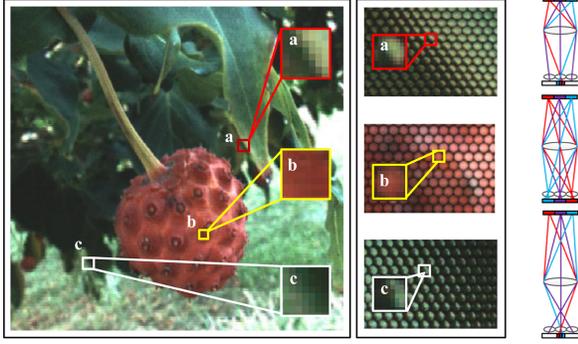


Fig. 1. The real light field images(raw images and images under each micro-lens) and the optical diagram. If the scene is behind the focusing plane, the micro-lens image is inverted compared with the scene view. If the scene is before the focusing plane, it is similar with the scene view. If the object is located at the focusing plane, it shows the consistent color.

which is still a time-consuming progress. Moreover, due to the heavy noises and spatial aliasing [14] in plenoptic light field images, the depth is more difficult to estimate.

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 [13], and then estimated in-focus regions in every image. Depth maps are 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 refocusing process and the in-focus region estimation needs to be calculated a lot of times to acquire the final depth map.

In contrast, our approach does not try to calculate the accurate depth map. Instead, we develop a simple method to utilize the structure information specific to light field image for saliency detection. The relative locations, with respect to the focused plane of the main lens, are calculated directly from the raw images and integrated to acquire the saliency results.

3. RELATIVE LOCATION EXTRACTION

In this paper, $L(x, y, u, v)$ is used 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. In this section, we try to extract 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, the optical diagram and the real images both show the relationship between images under micro-lens and scene views. Specifically, if the scene is before the focusing plane, the micro-lens image is similar with the scene view. On the contrary, it is inverted compared with the scene view. If the object is located at the focusing plane, the micro-lens image shows the consistent color of the same point in the scene.

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. In particular, we build two filters, foreground filter W_f and background filter W_b , to evaluate the possibilities of the points position. We modified a general linear filter as the popular bilateral [15] or guided filter [16] to acquire the location information. The proposed filter treats a view image I_v as a guidance image, the raw light field image I_r as an input image. The foreground filter W_f is constructed according to the view image I_v :

$$W_{ij}^f = \exp\left(-\frac{|I_v(p_j) - I_v(p_i)|^2}{2\delta^2}\right), \quad (1)$$

where $I_v(p_j)$ is pixels in a window around $I_v(p_i)$ in the view image. In this paper, we set the window size equal to the size of the micro-lens for convenient calculation in the filtering process. The experiments in the realistic scene prove that the window size is insensitive to most of the depth ranges. The background filter W^b is set as the transpose of the W^f accordingly.

The output image after filtering is then expressed as a weighted average of every single micro-lens image in the raw image I_r :

$$I(q_i) = \sum_j W_{ij}(I_v)(I_r(p_j) - I_r(p_i))^2, \quad (2)$$

where p_i is the center pixel of each micro-lens image and p_j is pixels in the same micro-lens. q_i is the corresponding output which has the same size as the view image. The two filters are applied to the raw images and the filtered images I^f and I^b are obtained.

3.2. Relative Location

As we analyzed before, if the point is behind the focusing plane, the micro-lens image is more similar with the inverted view image, which means I^f is larger than I^b . On the contrary, if the object is before the focusing plane, I^f is smaller 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 [16], and

then the relative location is defined as:

$$L = \frac{I^f - I^b}{I^f + I^b}, \quad (3)$$

where $L \gg 0$ indicates the point is more likely behind the focusing plane, $L \ll 0$ means it is before the focusing plane, and $L \approx 0$ means it is located on the focusing plane.

4. SALIENCY DETECTION

In this section, we show how to integrate the extracted relative location with the color information to obtain the final saliency map. First, we segment the reference image into a set of superpixels using mean-shift algorithm [17]. The relative location cues are normalized and then computed as the average value of all pixels within a region $l(p)$.

4.1. Background Selection

In the recent state-of-art saliency detection method, Zhu *et al.*[8] assumes that the salient objects are much less connected to image boundaries than background ones. In order to detect salient objects in complex backgrounds, we add the relative location cues in the modified method.

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

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

where the definition of $Len_{bnd}(p)$ is the length along the boundary and $Area(p)$ is a soft area that p belongs to, as defined in [8]. The difference with their work is that the $d(p, q)$ used in $Len_{bnd}(p)$ and $Area(p)$ does not only consider the color information, but also fuse the location information, which 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. Then the possibility of the background is defined as:

$$\omega_i^{bg} = 1 - \exp\left(-\frac{BndCon^2(p_i)}{2\sigma_{bndCon}^2}\right), \quad (5)$$

where σ_{bndCon} is used for adjusting boundary connectivities.

4.2. Contrast Selection

Similar with [8], we defined the contrasts of objects as:

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

where $l_{Area}(p)$ is a added term for averaging locations in the soft area $Area(p)$, which has a large value when the area is close to the camera. The other definitions can be found in [8].

Most RGBD saliency detection work defines that objects close to the camera are more likely to be salient. This assumption is partly correct except in two scenes. Firstly, the overall location of the object is close to the camera but is connected to the image boundaries. Secondly, the depth of the object is changing sharply, *e.g.* the ground or the flat desktop. Due to the added relative location cues, the above problems can be easily solved. If the overall location of one object is close to the camera and the boundary connectivity is large, the ω_i^{bg} will be large. Besides, if the depth of the object is changing sharply, l_{Area} is averaged to be relatively lower than the other objects. As a result, only objects which are relatively close to the camera and also far away from the image boundaries are set at a large contrast.

Finally, the saliency map is obtained by minimizing the cost function using least-square:

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

where $\omega_{i,j}$ is the smoothness part and can be found in [8].

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 saliency detection methods designed for light field image (LF [6]), RGBD images (ACD [7], LS [1]) and traditional RGB images (RB [8], BL [18], DSR [19], GS [20], MR [21], SF [22]). Our experimental results are evaluated with 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 salience detection are calculated using the depth from focusness method [23], which are also released in dataset [6].

The visual examples are shown in Fig. 2. We can observe that the relative location cues are able to distinguish the outstanding objects clearly and highlight the salient parts. Compared with LF [6], the salient 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 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 slightly higher recall rate compared with using the relative location cue because of the more precise depth information. However, it needs more complex calculation and we can choose the proper method according to the system requirement.

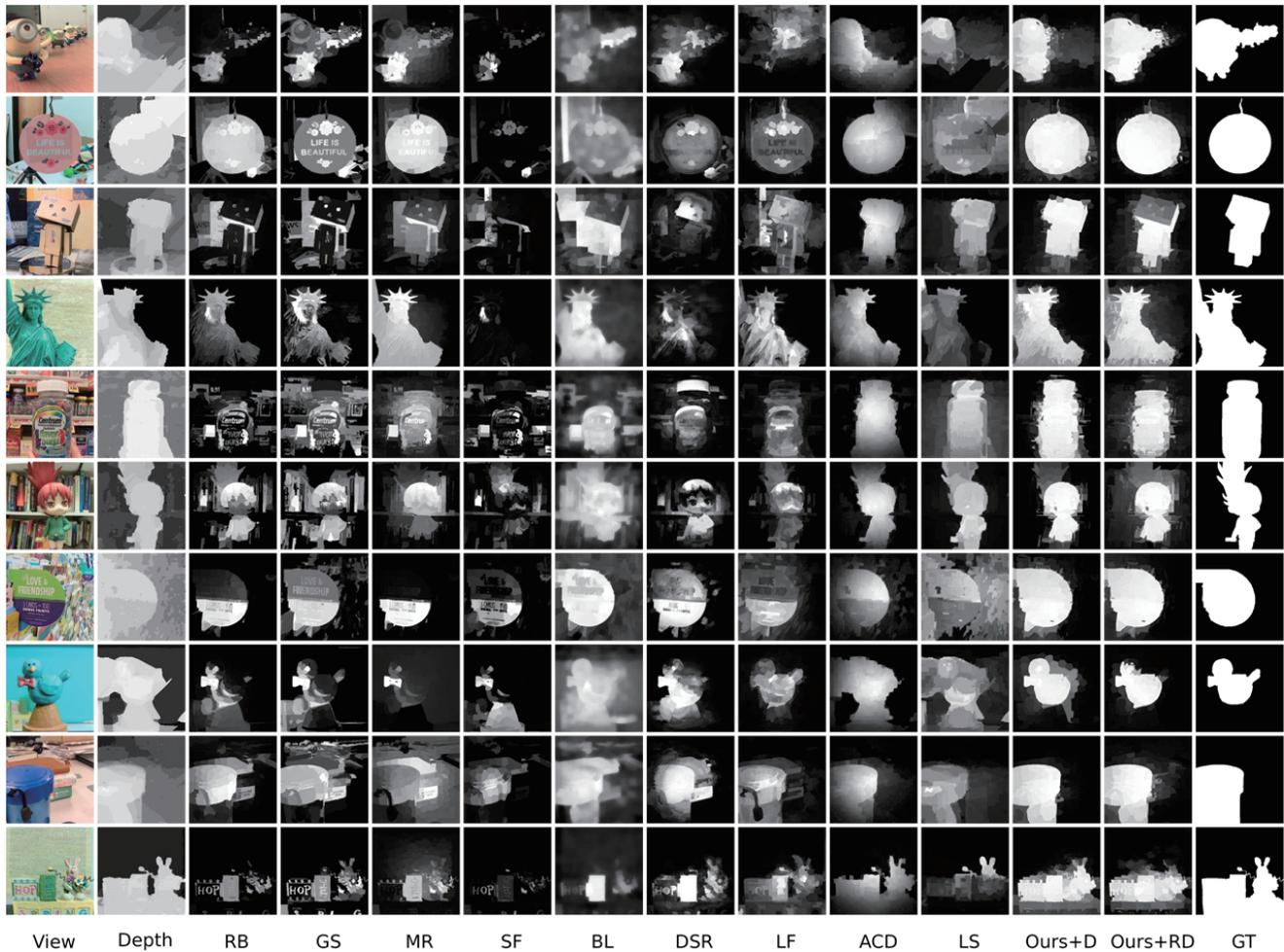


Fig. 2. Precision recall curve comparison with state-of-the-art methods.

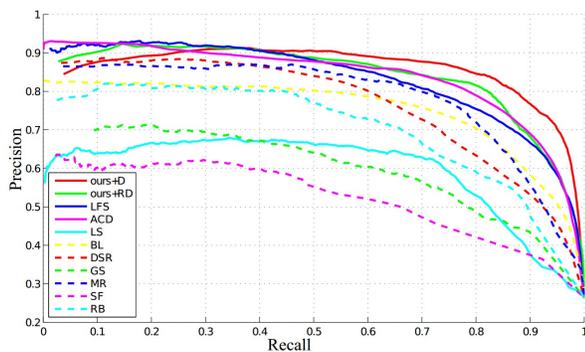


Fig. 3. Comparison of our saliency maps with state-of-the-art methods.

6. CONCLUSION

Taking into account the special structure of the light field images, we propose a novel relative location cues to extract the salient parts of an image. The relative location is calculat-

ed on the raw images, which is simple and effective. Based on the locations with respect to the focused plane, we can extract the salient regions using a modified saliency detection method. Compared with the state-of-the-art methods, the proposed method is able to detect saliency 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, "Rgb-d 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] 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.
- [10] 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.
- [11] 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*.
- [12] Gengliang Zhu Hao Sheng, Shuo Zhang and Zhang Xiong, "Guided integral filter for light field stereo matching," in *IEEE International Conference on Image Processing (ICIP), 2015*, p. 852856.
- [13] Marc Levoy, "Light fields and computational imaging," *IEEE Computer*, vol. 39, no. 8, pp. 46–55, 2006.
- [14] 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.
- [15] 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.
- [16] Kaiming He, Jian Sun, and Xiaoou Tang, "Guided image filtering," in *European Conference on Computer Vision (ECCV)*, pp. 1–14. Springer, 2010.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] Yichen Wei, Fang Wen, Wangjiang Zhu, and Jian Sun, "Geodesic saliency using background priors," in *Computer Vision—ECCV 2012*, pp. 29–42. Springer, 2012.
- [21] 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.
- [22] 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.
- [23] 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.