Cellular Automata Based on Occlusion Relationship for Saliency Detection

Hao Sheng[1,2(✉)], Weichao Feng[1], and Shuo Zhang[1]

[1] State Key Laboratory of Software Development Environment, School of Computer Science and Engineering, Beihang University, Beijing, China
{shenghao,fengwc,zhangshuo}@buaa.edu.cn
[2] Research Institute of Beihang University in Shenzhen, Shenzhen 518057, People's Republic of China

Abstract. Different from the traditional images, 4D light field images contain the scene structure information and have been proved that can better obtain the saliency. Instead of estimating depth or using the unique refocusing capability, we proposed to obtain the occlusion relationship from the raw image to calculate saliency detection. The occlusion relationship is calculated using the Epipolar Plane Image (EPI) from the raw light field image which can distinguish a region is most likely a foreground or background. By analyzing the occlusion relationship in the scene, true edges of objects can be selected from the surface textures of objects, which is effective to segment the object completely. Moreover, we assume that objects which are non-occluded are more likely to be the foreground and objects that are occluded by lots of objects are background. Then the occlusion relationship is integrated into a modified saliency detection framework to obtain the saliency regions. Experiment results demonstrate that the occlusion relationship can help to improve the saliency detection accuracy, and the proposed method achieves significantly higher accuracy and robustness in comparison with state-of-the-art light field saliency detection methods.

Keywords: Light field · Saliency detection · Occlusion relationship · Raw image · EPI image

1 Introduction

Saliency detection has been an important research direction in computer vision which is aiming at locating salient pixels or regions in an image. Precise and trustworthy saliency detection plays an important preprocessing role in many computer vision tasks, including image segmentation [1], object detection [2], image thumbnailing [3] and retargeting [4]. Due to the absence of the high-level knowledge, almost all the 2D methods based on traditional images rely on assumptions that appearance contrasts between objects and their surrounding regions are high. This assumptions are called contrast prior and are used in many saliency detection methods [5–10]. Besides contrast prior, boundary prior [8]

F. Lehner and N. Fteimi (Eds.): KSEM 2016, LNAI 9983, pp. 28–39, 2016.
DOI: 10.1007/978-3-319-47650-6_3

has been used in many approaches [6,8,9]. Image boundary regions are mostly identified as backgrounds and boundary priors are further used to enhance the saliency detection. Although many methods suggest that the boundary prior is effective, it has its own drawbacks.

We can't easily assume that the salient object is not in the boundary and the background is in the boundary. In many cases, this assumption is wrong. In this paper, a more reliable prior, which consider the occlusion information, is introduced to obtain the background and foreground measurement.

Saliency detection for light field images is first proposed in [9], which has proved that light field image can obtain more information than the image captured by traditional cameras. Traditional cameras only capture color (RGB) information, whereas the light field camera has capability of refocusing and obtain the depth of the scene. In this paper, we explore the salient object detection problem by using a different input: the raw picture of the Lytro camera. Lytro camera is a light field camera which mounts a lenslet array in front of the sensor. It is the first commercial light field cameras and has $328 * 328$ spatial resolution and $10 * 10$ angular resolution. A light field image can be viewed as an array of images captured by a camera array. Specifically, it captures the scene from different continues angels and we can obtain the parallax image. Like the binocular images and RGBD images, parallax images can calculate the depth information. However, the disparity estimation from parallax images has been a difficult problem for a long time and various advanced methods have been proposed. Moreover, light field images captured by Lytro camera are much more difficult to obtain the accurate depth because the microlens are closely arranged and the parallax images have a very small difference. Although the light field cameras are difficult to obtain the accurate depth, it can benefit saliency detection in numbers of ways, like the capability of post-capture refocusing [11].

In this paper, we obtain a novel information from light field image called occlusion relationship. The occlusion relationship is calculated using the Epipolar Plane Image (EPI) from the raw light field image. Then we try to utilize the occlusion relationship as a new feature to judge the foreground regions and background regions. In other words, we utilize the occlusion relationship to obtain a more reliable background measure instead of the boundary. Moreover, we modified the cellular automata methods [12] and combined the color priors and background priors to calculate the salience object with our occlusion relationship features from light field images and achieve comparable results with the state-of-the-art methods.

2 Related Work

Saliency detection can be categorized into two categories, bottom-up approaches [13–16] and top-down approaches. Bottom-up approaches usually exploit low-level cues such as feature, colors and spatial distances to construct saliency maps. Itti [17] propose a saliency model based on the neural network which integrates three feature. Harel [18] propose a graph-based saliency measure. In [19–21], a saliency models based on Bayesian has been proposed which

is exploiting low and mid level cues. Sun et al. [22] improve the Xie's model by introducing boundary and soft segmentation. Perrazzi [23] propose a contrast-based saliency filter and calculate saliency detection by uniqueness and spatial distribution of the patch on the image. Shen and Wu [24] combine lower-level feature and higher-level feature to construct a model based on the low-rank matrix recovery. Cheng et al. [5] utilize a soft abstraction method to remove the image useless details and calculate salient regions. Goferman et al. [25] propose a context-aware saliency algorithm to detect the image regions which represent the scene based on the four principles of the human's visual attention. [26] compute the contrast of the center and surround distribution of features which is based on the Kullback-Leibler divergence for salient object detection. In [27], a graph-based bottom-up method is proposed using manifold ranking. Zhu [28] construct a saliency object detection method based on boundary connectivity. Compared to bottom-up approaches, top-down approaches are task-driven and require supervised learning with manually labeled ground truth. Zhang [29] construct a Bayesian-based top-down methods. Judd [30] learned a top-down saliency model object detectors such as faces, humans, animals, and text. [31] propose a saliency model by jointly learning a conditional Random Field and a dictionary. The bottom-up approaches of saliency detection are data-driven and usually based on low-level visual information and more interesting in detection the local feature rather than a global feature. On the other hands, top and down approaches of saliency detection are interesting in detect object easy to obtain visual attention like human face and body. Consider the time complexity problem, bottom-up approaches usually efficient than top-down approaches. Our approach belongs to bottom-up approaches where we add an occlusion cues. Early saliency detection method exploits center priors which expected to exhibit a high contrast in the center. Koch and Itti [17] are the first to use the center prior to detection saliency. After that, many methods computer the center prior locally or globally. Local methods computer the difference within some nearby pixels or super pixels. Global methods computer the difference into the whole images by computer the power spectrum [32], color histogram [33], and so on. Even if the use of center priors are proved effective, Wei [8] use background priors to calculate saliency detection and have been proved that background priors are also effective methods. Zhu [28] observed the boundary connectivity as a robust background measure and calculate the relationship between foreground and background. In most approaches, they use the color character to judge foreground or background. However, most images in the LF have similar color so that these approaches can easily fail. Our approach calculates each EPI image to obtain the occlusion relationship and combine the color contrast, background prior to obtaining the saliency detection.

3 Occlusion Relationship Analysis Algorithm

In this paper, we present the 4D light field with $L(s, t, u, v)$ representation where (s, t) is the coordinate of the image in different views and (u, v) is the coordinate

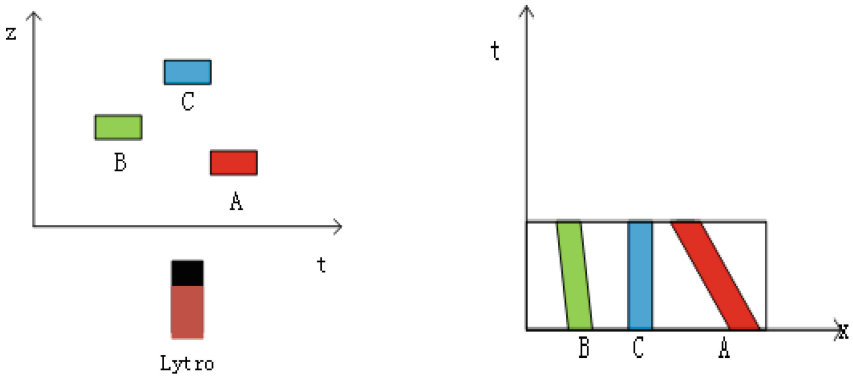


Fig. 1. Principle chart: the slope of the mutual occluded objects is different, the larger slope is occluded by the smaller slope

of the main lens plane. ST plane and UV plane are two parallel planes and $L(s,t,u,v)$ represent the intersection of incoming light from a different view perspective. Using $L(s,t,u,v)$ representation is time-saving in computing raw data and occlusion relation.

3.1 Hierarchical Relationship Model

Epipolar Plane Image (EPI) is constructed by collating multiple images taken from the equidistant location along a line. EPI slice can be taken from parallax image in equidistant location along a line. The point in parallax image maps to a straight line in EPI slice. The slope of EPI line is proportional to its depth in real space as Fig. 1 and adjacent EPI lines with different slope has an occlusion relationship in real space. In Fig. 1, there are three objects A, B, C in front of the Lytro camera, object A is the closest object to the camera which is more possible to be the salience object and C is the farthest object to the camera which has a little possible to be the salience object. Object A and B are closer to the camera than C, so object C is occluded by the object A and B. As a result, object A and B may be more salience because they shelter C. In EPI slice, there are three trails and the largest slope represented to C and A has the smallest slope. In other words, the smaller slope of the trail represents that the point is closer to the camera and the larger slope of the trail represent that the point is far from the camera. So, the object has smaller slope occluded the object which has larger slope.

Based on the observations, we construct a specific feature to present whether a pixel is occluded by other or shelter others. After obtained the EPI image, we first apply Canny edge detection on the EPI slice. Then we curve fit all the point on the line and obtain the slope of each line. Because of the texture, adjacent trails may belong to the same object, so we fuse the adjacent trails which have the same color feature as follows:

(a) RGB image (b) occlusion relationship

Fig. 2. The scene image and the result of occlusion relationship image. The red point represents the occlusion point and the green represents the occluded point (Color figure online)

$$t_i = ||c_i - c_j||, \tag{1}$$

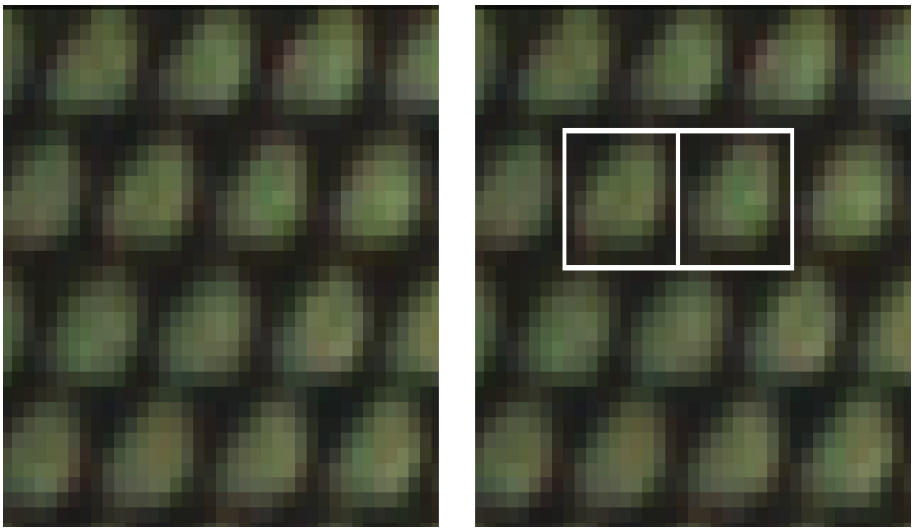
where $||c_i - c_j||$ is the Euclidean Distance between the superpixel i and j in CIELAB color space. We consider the superpixel i and j belongs to the same object when $t_i < 10$. Then we calculate the adjacent lines in the EPI lines using their slope and the equation as:

$$T_i = s_i - s_j, \tag{2}$$

where s_i is the slope the line i. If the $T_i < -0.15$, we considered that the line i occluded line $i+1$. using this equation, we can obtain the occlusion relationship. Then collect all the relationship of each EPI slice and obtain the all occlusion relationship of the origin image as Fig. 2.

3.2 Hierarchical Relationship Based on Lytro Camera

Images photoed by Lytro camera are stored in the .lfp file format as their RAW data and we can not handle to extract parallax images directly. So we must convert it into the files we can compute. In [34], we know the .lfp file format contains a lot of information such as focal length and gamma parameter in the file header. The RAW image file is a gray-scale image with BGGR Bayer pattern to store different RGB channels values. We first demosaic the RAW image and then estimate the rotation from the sensor and the micro-lens. After rotated the image, we need to obtain each micro-lens image and calculate the epipolar image. Due to the rotation, the total resolution of the image is less than $3280 * 3280$ and



(a) image after rotation (b) segment raw image

Fig. 3. Obtain the micro-len image

the total number of micro-lens is $328 * 328$, so some micro-lens are $10 * 10$ and some are less than $10 * 10$. We utilize the characteristics that there are a much darker line between the adjacent micro-lens. We obtain each micro-lens image using this method and then obtain the EPI image. The left image in Fig. 3 is the image after rotation and the right image extracts the middle two micro-lens image.

Because of each miro-lens has about $10 * 10$ pixels, we can obtain about $10 * 10$ parallax pictures. When we use the 10 pictures in the x direction, we can obtain the EPI in x direction. Similarly, when we use the 10 pictures in the y direction, we can obtain the EPI in y direction. So we can obtain the occlusion relationship in both x and y direction. Traditional camera array can obtain occlusion relationship in only one direction and we can obtain the occlusion relationship in two directions easily.

3.3 Background Measure

We observe that foreground and background regions in natural images and datasets are quite different. Foreground region tends to be in a small depth and in the front of the image, on the contrary, background regions tend to be in a high depth value and far from the image. It is intuitive cognition to the foreground and background and we considered the occluded region as a background point and occlude another region as a foreground point. To accurately measure background regions and improve computational efficiency, an input image is segmented into N small superpixels by the simple linear iterative clustering (SLIC) algorithm [35]. The number background points and foreground points in each

superpixel are accumulated and the background measure is defined as: Where indicate the region belongs to a background and indicate the region belongs to the foreground.

4 Saliency Detection

In this section, we detect the salience region of the image using background measure combined with the RGB color information. We show how to use our information to obtain the final salience result and compare with another methods. The rest salience detection frame is based on the existed cellular automate method worked by [12] and we add the occlusion relationship cues into the method in the first section to obtain a better salience result. To improve computational efficiency, we segmented the input image into N small superpixels using simple linear iterative clustering(SLIC) [35] algorithm. We utilize the superpixel instead of pixels to calculate saliency.

In [12], background seeds are image boundary and the K-means algorithm is applied to divide it into 3 clusters based on the CIELAB color features. However, if the background is complex, we cannot effectively select enough background seeds and cannot obtain a reasonable result for global color distinctionm (GCD) maps. We utilize the background pixels as background seed and then apply the K-means algorithm to obtain the GCD map. GCD map is computed as: where is defined as in [12], after calculated global color distinction maps, we calculated global spatial distance (GSD) matrix.

Based on the GCD map and GSD matrix, the single-layer cellular automate (SCA) method is used to calculate the salient regions. Each cell denotes a superpixel and the saliency value of each superpixel denotes its state and it is discrete. Updating rule is the main point of the cellular automata and is defined as:

$$S^{t+1} = C^* \cdot S^t + (I - C^*) \cdot F^* \cdot S^t, \tag{3}$$

as in [12], the saliency map is computed after N1 times.

5 Experiment

We evaluate the proposed method on Light field (LF) datasets [9], which are captured by Lytro camera and include 100 light field images with the estimated depth map and salient region ground truth. We compare our method with state-of-the-art saliency detection methods, including LF [9], MR [27], GS [8], SF [23], wCtr [28]. LF [9] is based on the light field camera and MR [27] which is used on Manifold Ranking, GS [8] which is used the geodesic saliency using background priors, SF [23] which is used saliency filters to obtain the salient region and wCtr [28] which is used on the background priors are all based on the RGB space. In order to prove the effectiveness of the proposed occlusion relationship, we also evaluate the saliency results using depth cues instead of the occlusion relationship. We use depth map from select the most far from the screen which has more great depth value. We sort all the depth values and select

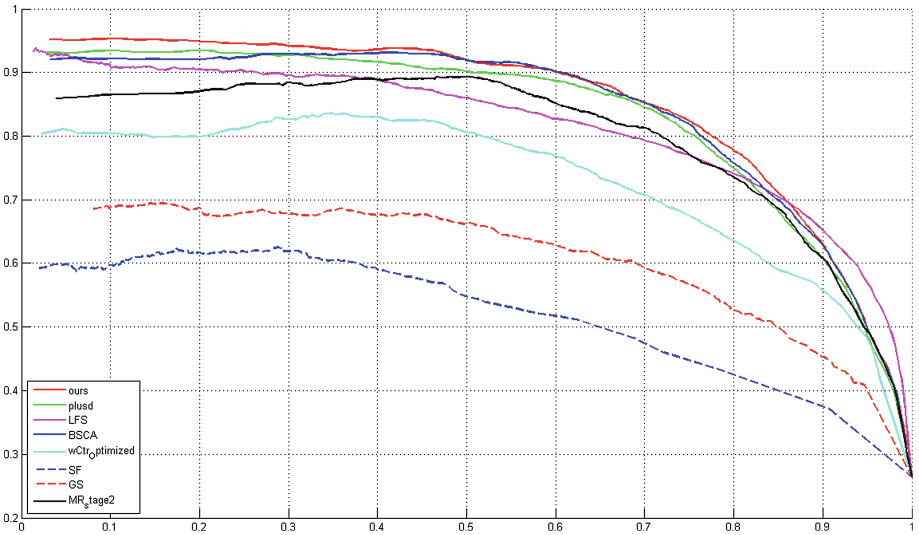


Fig. 4. Precision recall curve comparation with state-of-the-art methods

the maximum 20 % depth pixels as background points and use cellular automate to obtain saliency detection.

We evaluate our saliency result with occlusion relationship cues and depth cues to show the effectiveness of our light field cues. We can observe that the depth cues is useful in the saliency detection because the depth cues help the background priors more precisely. By contrast, the occlusion relationship is able to detect the background more precise than the depth cues. The reason is the occlusion relation can find the relatively background and is more precise than the background in the all image.

The visual examples are showed in Fig. 5, we can see that the occlusion relationship cues are able to improve the cellular automate method and highlight the salience parts.

We follow the canonical precision-recall curve (PRC) to evaluate the accuracy of the detected saliency on databases of different dimension. For details about PRC methods, we refer readers to [17] and the parameters setting in our implement is the same as [36]. Figure 4 shows the result of the PRC comparison architecture. Our method achieves the highest precision in most of the recall range on all datasets. Occlusion relationship and depth cues can improve the salience and our feature achieves a higher precision and recall rate compared with using depth cues. The reason is that the depth cues always located at a corner and our cues are around all the map.

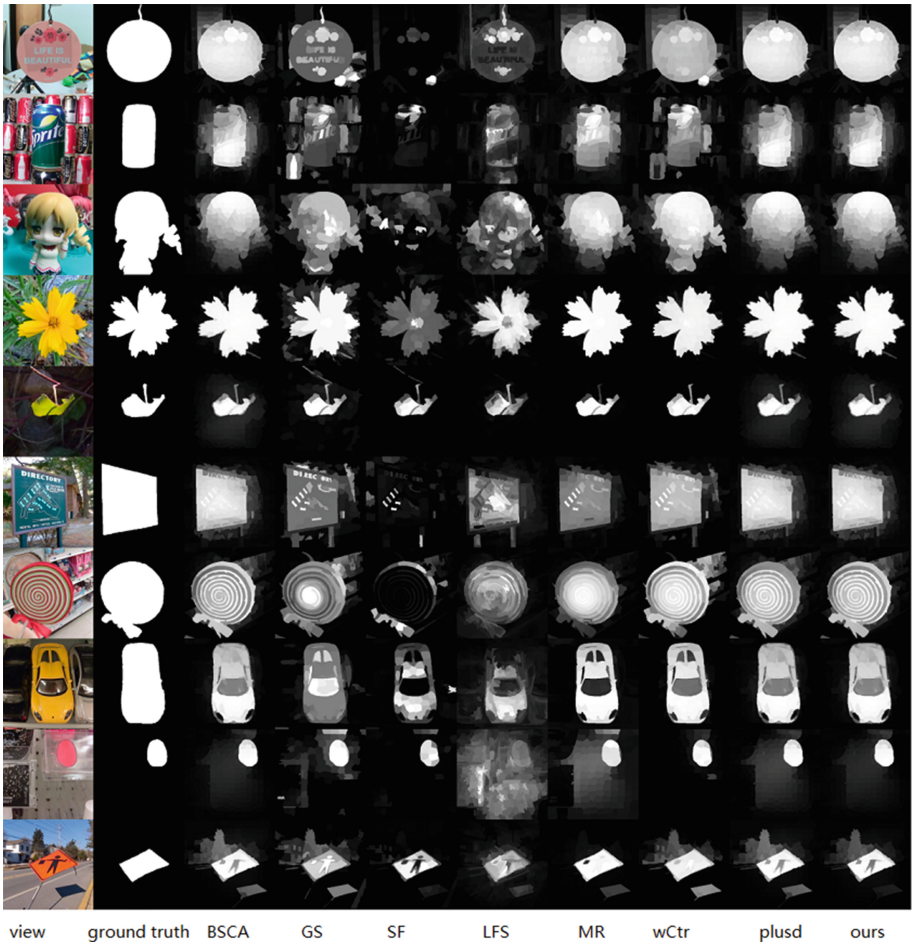


Fig. 5. Comparison of our saliency maps with state-of-the-art methods

6 Conclusion

We propose a bottom-up method to detect salient regions in light field images through cellular automata method, which incorporates occlusion relationship cues and background and foreground priors. The occlusion relationship is calculated on the raw images and is showed that it is simple and effective. Based on the EPI images, we can obtain the occlusion relationship and use this cues to calculate the background and foreground. The information then extracts the location of the possible background and foreground in the image. We evaluate the proposed algorithm on large datasets and demonstrate promising results with comparisons to several state-of-the-art methods. Furthermore, the proposed algorithm is efficient. Our future work will focus on integration of multiple features with applications to other vision problems in light field.

References

1. Rother, C., Kolmogorov, V., Blake, A.: GrabCut: interactive foreground extraction using iterated graph cuts. ACM Trans. Graph. **23**(3), 307–312 (2004)
2. Borji, A., Sihite, D.N., Itti, L.: Salient object detection: a benchmark. IEEE Trans. Image Process. **24**(12), 414–429 (2015)
3. Sun, J., Ling, H.: Scale and object aware image retargeting for thumbnail browsing. 2011 IEEE International Conference on Computer Vision (ICCV), pp. 1511–1518. IEEE (2011)
4. Ding, Y., Xiao, J., Yu, J.: Importance filtering for image retargeting. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 89–96. IEEE (2011)
5. Cheng, M.-M., Warrell, J., Lin, W.-Y., Zheng, S., Vineet, V., Crook, N.: Efficient salient region detection with soft image abstraction. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1529–1536 (2013)
6. Jiang, H., Wang, J., Yuan, Z., Yang, W., Zheng, N., Li, S.: Salient object detection: a discriminative regional feature integration approach. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2083–2090 (2013)
7. Jiang, Z., Davis, L.S.: Submodular salient region detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2043–2050 (2013)
8. Wei, Y., Wen, F., Zhu, W., Sun, J.: Geodesic saliency using background priors. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7574, pp. 29–42. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33712-3_3
9. Li, N., Ye, J., Ji, Y., Ling, H., Yu, J.: Saliency detection on light field. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2806–2813 (2014)
10. Yan, Q., Xu, L., Shi, J., Jia, J.: Hierarchical saliency detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1155–1162 (2013)
11. Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., Hanrahan, P.: Light field photography with a hand-held plenoptic camera. Computer Science Technical Report CSTR, vol. 2, no. 11, pp. 1–11 (2005)
12. Qin, Y., Lu, H., Xu, Y., Wang, H.: Saliency detection via cellular automata. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 110–119 (2015)
13. Achanta, R., Estrada, F., Wils, P., Süsstrunk, S.: Salient region detection and segmentation. In: Gasteratos, A., Vincze, M., Tsotsos, J.K. (eds.) ICVS 2008. LNCS, vol. 5008, pp. 66–75. Springer, Heidelberg (2008). doi:10.1007/978-3-540-79547-6_7
14. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 1597–1604. IEEE (2009)
15. Bruce, N., Tsotsos, J.: Saliency based on information maximization. In: Advances in Neural Information Processing Systems, pp. 155–162 (2005)
16. Zhai, Y., Shah, M.: Visual attention detection in video sequences using spatiotemporal cues. In: Proceedings of the 14th ACM International Conference on Multimedia, pp. 815–824. ACM (2006)
17. Itti, L., Koch, C., Niebur, E., et al.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. **20**(11), 1254–1259 (1998)

18. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: Advances in Neural Information Processing Systems, pp. 545–552 (2006)
19. Rahtu, E., Kannala, J., Salo, M., Heikkilä, J.: Segmenting salient objects from images and videos. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6315, pp. 366–379. Springer, Heidelberg (2010). doi:10.1007/978-3-642-15555-0_27
20. Xie, Y., Lu, H.: Visual saliency detection based on bayesian model. In: 2011 18th IEEE International Conference on Image Processing (ICIP), pp. 645–648. IEEE (2011)
21. Xie, Y., Huchuan, L., Yang, M.-H.: Bayesian saliency via low and mid level cues. IEEE Trans. Image Process. **22**(5), 1689–1698 (2013)
22. Sun, J., Lu, H., Li, S.: Saliency detection based on integration of boundary and soft-segmentation. In: 2012 19th IEEE International Conference on Image Processing, pp. 1085–1088. IEEE (2012)
23. Perazzi, F., Krähenbühl, P., Pritch, Y., Hornung, A.: Saliency filters: contrast based filtering for salient region detection. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 733–740. IEEE (2012)
24. Shen, X., Wu, Y.: A unified approach to salient object detection via low rank matrix recovery. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 853–860. IEEE (2012)
25. Goferman, S., Zelnik-Manor, L., Tal, A.: Context-aware saliency detection. IEEE Trans. Pattern Anal. Mach. Intell. **34**(10), 1915–1926 (2012)
26. Klein, D.A., Frintrop, S.: Center-surround divergence of feature statistics for salient object detection. In: 2011 International Conference on Computer Vision, pp. 2214–2219. IEEE (2011)
27. Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M.-H.: Saliency detection via graph-based manifold ranking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3166–3173 (2013)
28. Zhang, X., Wang, Z., Yan, C., Zou, H., Peng, Q., Jiang, X., Dan, W.U.: Animal Nutrition Institute, and Sichuan Agricultural University, "Saliency optimization from robust background detection". In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2814–2821 (2014)
29. Zhang, L., Tong, M.H., Marks, T.K., Shan, H., Cottrell, G.W.: Sun: a bayesian framework for saliency using natural statistics. J. Vis. **8**(7), 32–32 (2008)
30. Judd, T., Ehinger, K., Durand, F., Torralba, A.: Learning to predict where humans look, vol. 30, no. 2, pp. 2106–2113 (2009)
31. Yang, J.: Top-down visual saliency via joint CRF and dictionary learning. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Proceedings/CVPR. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 157, no. 10, pp. 1–1 (2012)
32. Hou, X., Zhang, L.: Saliency detection: a spectral residual approach. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
33. Cheng, M., Zhang, G., Mitra, N.J., Huang, X., Hu, S.: Global contrast based salient region detection. IEEE Trans. Pattern Anal. Mach. Intell. **37**(3), 409–416 (2015)
34. Cho, D., Lee, M., Kim, S., Tai, Y.W.: Modeling the calibration pipeline of the Lytro camera for high quality light-field image reconstruction. In: Proceedings of the 2013 IEEE International Conference on Computer Vision, pp. 3280–3287 (2013)

35. Radhakrishna, A., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: Slic superpixels, Dept. School Comput. Commun. Sci., EPFL, Lausanne, Switzerland. Technical report 149300 (2010)
36. Cheng, M.-M., Mitra, N.J., Huang, X., Torr, P.H.S., Hu, S.-M.: Global contrast based salient region detection. IEEE Trans. Pattern Anal. Mach. Intell. **37**(3), 569–582 (2015)