Saliency Detection with a Deeper Investigation of Light Field

Jun Zhang[†], Meng Wang[†], Jun Gao[†], Yi Wang[†], Xudong Zhang[†], and Xindong Wu[†,‡]

[†] School of Computer Science and Information Engineering, Hefei University of Technology, China
[‡] Department of Computer Science, University of Vermont, USA
zhangjun@hfut.edu.cn, eric.mengwang@gmail.com, gaojun@hfut.edu.cn
wangyi916@mail.hfut.edu.cn, xudong@hfut.edu.cn, xwu@uvm.edu

Abstract

Although the *light field* has been recently recognized helpful in saliency detection, it is not comprehensively explored yet. In this work, we propose a new saliency detection model with light field data. The idea behind the proposed model originates from the following observations. (1) People can distinguish regions at different depth levels via adjusting the focus of eyes. Similarly, a light field image can generate a set of focal slices focusing at different depth levels, which suggests that a background can be weighted by selecting the corresponding slice. We show that background priors encoded by light field focusness have advantages in eliminating background distraction and enhancing the saliency by weighting the light field contrast. (2) Regions at closer depth ranges tend to be salient, while far in the distance mostly belong to the backgrounds. We show that foreground objects can be easily separated from similar or cluttered backgrounds by exploiting their light field depth. Extensive evaluations on the recently introduced Light Field Saliency Dataset (*LFSD*) [Li *et al.*, 2014], including studies of different light field cues and comparisons with Li *et al.*'s method (the only reported light field saliency detection approach to our knowledge) and the 2D/3D state-of-the-art approaches extended with light field depth/focusness information, show that the investigated light field properties are complementary with each other and lead to improvements on 2D/3D models, and our approach produces superior results in comparison with the state-of-the-art.

1 Introduction

Saliency detection, aiming at identifying the most salient regions or objects that most attract the viewers' visual attention in a scene, has become a popular area in computer vision. It plays an important role in recognition [Rutishauser *et al.*, 2004; Han and Vasconcelos, 2014], image segmentation [Goferman *et al.*, 2012], and visual tracking [Mahadevan and Vasconcelos, 2012].

Depending on how data are used, saliency detection can be classified into three main categories: 2D, 3D and light field saliency. The most common approach is to apply 2D salient cues such as intensity and color in contrast-based data-driven bottom-up saliency frameworks [Itti and Koch, 2001; Cheng *et al.*, 2011; Perazzi *et al.*, 2012; Jiang *et al.*, 2013] owing to the observation that human vision is either particularly sensitive to high-contrast stimuli [Reynolds and Desimone, 2003], or incorporated with object/background priors in context-dependent top-down mechanisms [Wei *et al.*, 2012; Yang *et al.*, 2013; Zhu *et al.*, 2014]. A comprehensive study on existing 2D saliency detection approaches is conducted by Borji *et al.* [2012]. Although these 2D models are based on visual attention mechanisms especially rooted in an early visual process in primary visual cortex (area *V1*) [Koch and Ullman, 1985; Itti and Koch, 2001; Li, 2002], most methods ignore important aspects of eye movements such as attention shifting across depth planes [Jansen *et al.*, 2009].

With the availability of commercial 3D cameras such as *Kinect* [Zhang, 2012], another approach to saliency detection involves the saliency computation of RGBD images [Zhang *et al.*, 2010; Lang *et al.*, 2012; Desingh *et al.*, 2013; Ciptadi *et al.*, 2013; Peng *et al.*, 2014; Ju *et al.*, 2014]. In these methods, depth, as one of the feature dimensions, is more directly bound to objects thus beneficial for saliency analysis. However, most current methods require high-quality depth maps and ignore the relations between depth and appearance cues [Peng *et al.*, 2014].

In recent years, the *light field* has opened up a new research area with the development of digital photography. Using consumer light field cameras such as *Lytro* [Ng *et al.*, 2005] and *Raytrix* [Lumsdaine and Georgiev, 2009], we can simultaneously capture the total amount of light intensity and the direction of each ray from incoming light in a single exposure. Therefore, a light field can be represented as a 4D function of light rays in terms of their (2D spatial) positions and (2D angular) directions [Adelson and Wang, 1992; Gortler *et al.*, 1996]. These information items can be converted into various interesting 2D images (*e.g.*, focal slices, depth maps and all-focus images) through rendering and refocusing techniques [Ng *et al.*, 2005; Tao *et al.*, 2013]. For example, a light field image can generate a set of focal slices focusing at different depth levels [Ng *et al.*, 2005] (see Figure 1(a) and 1(b) for examples), which suggests that one can
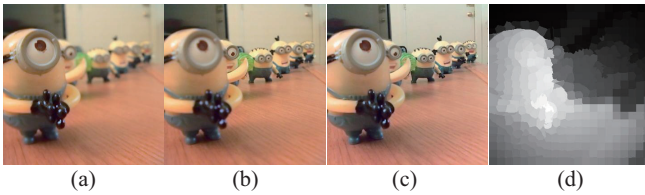
| (a) | (b) | (c) | (d) |

Figure 1: Focal stack, all-focus image and depth map obtained from a Lytro camera. (a) and (b) are two slices that focus at different depth levels (left is focused on the first minion and right is focused on the second minion); (c) all-focus image; (d) depth map.

determine background regions by selecting the slice focused on the background. Then the corresponding focusness maps can be computed to separate in-focus and our-of-focus regions so as to identify salient regions. An all-focus image (Figure 1(c)), containing a series of images captured at different focal planes, provides the sharpest pixels and can be approximately recovered by depth invariant blurs from the focal stack images [Nagahara *et al.*, 2008] or the depth of field (*DOF*)-dependent rendering algorithm [Zhang *et al.*, 2014]. In addition, the depth of each ray of light recorded in the sensor can be estimated by measuring pixels in the focus [Tao *et al.*, 2013], as shown in Figure 1(d). Regions at closer depth ranges tend to be salient, while far in the distance mostly belong to backgrounds.

Li *et al.* [2014] pioneered the idea of light field for resolving traditionally challenging problems in saliency detection. They proposed a new saliency detection method tailored for light field by combining foreground and background cues generated from focal slices. In addition, they collected the first challenging dataset of 100 light fields using Lytro camera, *i.e.*, the Light Field Saliency Dataset (LFSD). They demonstrated that the light field can greatly improve the accuracy of saliency detection within challenging scenarios. Although their work aims to explore the role of light field in saliency detection by using focusness to facilitate the saliency estimation, it is still at the initial stage of exploration and has some limitations: (1) the focusness and objectness are calculated to select foreground saliency candidates, which inevitably ignores the explicit use of depth data associated with salient regions/objects, and (2) the performance of existing 2D/3D saliency detection approaches and their extended versions with light field cues on the LFSD dataset is not well explored.

In this paper, we make an attempt to address saliency detection by the use of light field properties in the following aspects: **(1)** we generate focusness maps from focal slices at different depth levels via invariant blurs [Shi *et al.*, 2014] and introduce a light field depth cue into saliency contrast computation within a L_2-norm metric, **(2)** to facilitate the saliency estimation, we compute the background prior on the focusness map of a selected focal slice and incorporate it with the location prior, **(3)** we extend the 2D/3D state-of-the-art saliency detection approaches with our proposed light field depth contrast and focusness-based background priors, and

show the effectiveness and superiority of light field properties, and **(4)** comparing with [Li *et al.*, 2014], our approach increases the performance by 4%–7% on the LFSD dataset in terms of different quantitative measurements.

2 Related Work

2.1 2D Saliency

Results from many existing visual saliency approaches [Itti and Koch, 2001; Cheng *et al.*, 2011; Perazzi *et al.*, 2012; Jiang *et al.*, 2013] indicate that the contrast is the most influential factor in the bottom-up visual saliency. For examples, Itti *et al.* [1998] proposed a saliency model that computes the local contrast from color, intensity and orientation. Cheng *et al.* [2011] defined the contrast by computing dissimilarities among color histogram bins of all image regions. To take global spatial relationships into account, Perazzi *et al.* [2012] considered saliency estimation as two Gaussian filters performing on region uniqueness and spatial distribution respectively. Other global methods such as appearance reconstruction [Li *et al.*, 2013] and fully connected MRF [Jia and Han, 2013] are recently proposed to uniformly identify salient objects. Recently, background priors are incorporated into proposed methods to reduce the distraction of salient regions from backgrounds [Yang *et al.*, 2013; Wei *et al.*, 2012; Zhu *et al.*, 2014]. Such approaches are generally based on some specific assumptions that all the image patches heavily connected to backgrounds and image boundaries belong to backgrounds [Yang *et al.*, 2013; Zhu *et al.*, 2014], are very large and homogeneous [Wei *et al.*, 2012]. Additionally, Jiang *et al.* [2013] showed that the focus blur in 2D images can improve fixation predictions. Despite many recent improvements, the generations of accurate saliency maps are still very difficult in some challenging scenes, such as cluttered backgrounds, textured objects, and similar colors between salient objects and their surroundings, *etc* (see Figure 7 for some examples).

2.2 3D Saliency

Besides 2D information, several studies have exploited the depth cue in saliency analysis [Zhang *et al.*, 2010; Lang *et al.*, 2012; Desingh *et al.*, 2013; Peng *et al.*, 2014; Ciptadi *et al.*, 2013; Ju *et al.*, 2014]. Specifically, Zhang *et al.* [2010] designed a stereoscopic visual attention algorithm for 3D video based on multiple perceptual stimuli. Desingh *et al.* [2013] fused appearance and depth cues by using non-linear support vector regression. Ciptadi *et al.* [2013] demonstrated the effectiveness of 3D layout and shape features from depth images for calculating more informative saliency map. Ju *et al.* [2014] proposed a depth saliency method based on anisotropic center-surround difference, and used depth and location priors to refine the saliency map. Recently, a large-scale RGBD salient object detection benchmark is built up with unified evaluation metrics and a multi-stage saliency estimation algorithm is proposed to combine depth and appearance cues [Peng *et al.*, 2014]. The above approaches demonstrate the effectiveness of the depth in saliency detection, while their performance is highly dependent on the quality
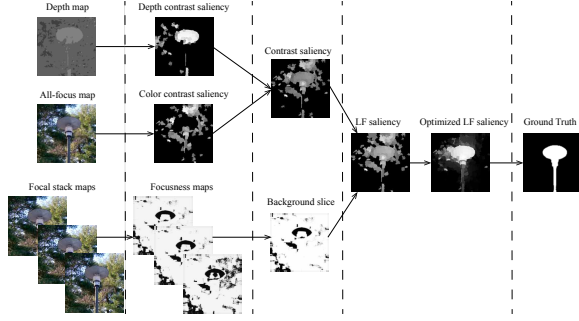
Figure 2: Pipeline of our proposed approach for light field saliency detection.

of depth estimation. All these methods may fail when salient objects cannot be distinguished at the depth level.

2.3 Light Field Saliency

To the best of our knowledge, Li *et al.*'s work [2014] is the first saliency detection method by using light field data that shows that light field can greatly improve the accuracy of saliency detection. They used focusness priors to extract background information and computed the contrast-based saliency between background and non-background regions. In addition, the objectness is computed as the weight for combining contrast-/focusness-based saliency candidates to generate the final saliency map.

3 Approach

Figure 2 shows the pipeline of our approach, and the details are described in the following sections.

3.1 Light Field Contrast-based Saliency

We build the contrast-based saliency based on the light field depth and color from the all-focus image. We employ Simple Linear Iterative Clustering (SLIC) algorithm [Achanta *et al.*, 2012] to segment an all-focus image into a set of nearly regular super-pixels, which can preserve edge consistency and yield compact super-pixels. We define the contrast saliency $\mathbf{S}(p_i)$ for super-pixel p_i as:

$$\mathbf{S}(p_i) = \sum_{j=1}^{N} \mathbf{W}_{pos}(p_i)||U_{fea}(p_i) - U_{fea}(p_j)|| \quad (1)$$

$$\mathbf{W}_{pos}(p_i, p_j) = exp(-\frac{||U_{pos}^*(p_i) - U_{pos}^*(p_j)||^2}{2\sigma_w^2}) \quad (2)$$

where N is the total number of super-pixels and we found that 300 super-pixels are enough to obtain high performance for saliency detection. $U_{fea}(p_i)$ and $U_{fea}(p_j)$ are average feature (depth or color) values of super-pixels p_i and p_j. $\mathbf{W}_{pos}(p_i, p_j)$ is the spatial weight factor for controlling the

pair-wise distance of super-pixels, that is, closer regions or similar colors would have higher contribution to the saliency. $||U_{pos}^*(p_i) - U_{pos}^*(p_j)||$ defines the L_2-norm distance of normalized average coordinates between pairs of super-pixels p_i and p_j and measures spatial relationships of super-pixels, σ_w is specified as 0.67 throughout our experiments.

Based on the above definitions, we denote the depth-induced contrast saliency from the light field as $\mathbf{S}_D(p_i)$, which is useful for separating the salient foreground from the similar or cluttered background (*e.g.*, the 1st example in Figure 4(c)). However, saliency detection with the depth contrast may fail when the background is close to the foreground object or the salient object is placed in the background, as shown in the 2nd example of Figure 4(c). We solve this issue by considering the color contrast as a complementary prior. Therefore, we compute the color contrast saliency $\mathbf{S}_C(p_i)$ in CIE LAB color space on the all-focus image.

Then, we combine the depth saliency and color saliency as:

$$\mathbf{S}^*(p_i) = \alpha \times \mathbf{S}_C(p_i) + \beta \times \mathbf{S}_D(p_i) \quad (3)$$

where α and β are two weight parameters for leveraging depth and color cues with $\beta = 1 - \alpha$. We empirically set them as $\alpha = 0.3$ and $\beta = 0.7$.

3.2 Background Priors Encoded by Focusness

Similar with [Li *et al.*, 2014], we select the background slice I_{bg} through analyzing focusness distributions of different focal slices $I_k, k = 1, ..., K$. More specifically, we compute the focusness map F_k for each focal slice using focusness detection technique [Shi *et al.*, 2014].

We compute the background likelihood score B_k for each slice I_k along $F_k(x)$ and $F_k(y)$ by U-shaped filtering, and choose the slice with the highest B_k as the background slice I_{bg},

$$F_{bg} = \arg\max_{k=1,...,K} B_k(F_k, u) \quad (4)$$

where $u = \frac{1}{\sqrt{1+(\frac{x}{\eta})^2}} + \frac{1}{\sqrt{1+(\frac{(w-x)}{\eta})^2}}$ is the 1D band-pass filtering function along the x axis, and $\eta = 28$ controls the bandwidth.

To enhance the saliency contrast, we compute the background probability Pb_{bg} on the focusness map F_{bg} through:

$$Pb_{bg}(p_i) = 1 - exp(-\frac{U_{bg}(p_i)^2}{2\sigma_{bg}^2} \cdot ||C - U_{pos}^*(p_i)||^2) \quad (5)$$

where $\sigma_{bg} = 1$, $U_{bg}(p_i)$ is the average value of super-pixel p_i on the focusness map F_{bg}. $||C - U_{pos}^*(p_i)||$ measures the spatial information of super-pixels related to the image center C, here, $U_{pos}^*(p_i)$ defines normalized average coordinates of super-pixel p_i. Therefore, regions that belong to the background have higher background probability Pb_{bg} on the focusness map.

3.3 Background Weighted Saliency Computation

We incorporate the background probability into the contrast saliency as follows:

$$\mathbf{S}_{lf}(p_i) = \sum_{j=1}^{N} \mathbf{S}^*(p_i) \cdot Pb_{bg}(p_j) \quad (6)$$

It can be seen that the saliency value of a foreground region is increased by multiplying a high Pb_{bg} from background regions. On the contrary, the saliency value of background regions is reduced by multiplying a small Pb_{bg} from foreground regions.

In order to obtain cleaner foreground objects, we applied saliency optimization algorithm [Zhu *et al.*, 2014] onto the above saliency map (Eq. 6). We found that the addition of this optimization procedure consistently increases the performance of the proposed approach by about 2% for MAE, 5% for F-measure, and 6% for AUC.

4 Experiments

We conduct extensive experiments to demonstrate the effectiveness and superiority of our proposed approach.

4.1 Experimental Setup

Dataset

Light Field Saliency Dataset (LFSD)[1] is the only reported dataset captured by *Lytro* camera for saliency analysis, which contains 100 light fields acquired in 60 indoor and 40 outdoor scenes. The light field data of a scene are composed of 5 types of image data – the raw light field data, a focal slice, an all-focus image derived from the focal stack, a rough depth map, and the corresponding 2D binary ground truth (GT).

Evaluation Measures

We use standard precision recall (PR) and receiver operating characteristic (ROC) curves for evaluations. When computing the overall quality on the whole dataset, we consider three metrics for determining the accuracy of saliency detection: F-measure, area under curve (AUC), and mean absolute error (MAE).

Given a saliency map, a PR curve is obtained by generating binary masks with a threshold $t \in [0, 255]$ and comparing these masks against the GT to obtain precision and recall rates. The PR curves are then averaged over the dataset. We compute F-measure as $F_\beta = \frac{(1+\beta^2)\cdot P\cdot R}{\beta^2\cdot P+R}$, and set $\beta^2 = 0.3$ to highlight precision [Achanta *et al.*, 2009]. The ROC curve can also be generated based on true and false positives obtained during the calculation of PR curve. MAE measures the average per-pixel difference between the binary GT and the saliency map, which is found complementary to PR curves [Perazzi *et al.*, 2012; Cheng *et al.*, 2013].

4.2 Results

Evaluating the Different Light Field Properties

To assess the different light field properties in the proposed approach, we show comparisons of the accuracy using different model components. Here we focus on different light field properties themselves regardless of saliency optimization [Zhu *et al.*, 2014], .

Figure 3 shows PR and ROC curves of saliency detection. It can be seen that light field properties complement each other and none of them alone suffices to achieve good results. When linearly combining the depth and color contrast
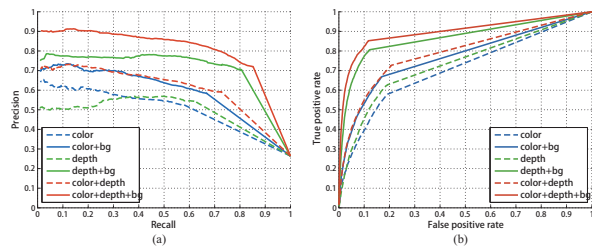
Figure 3: Quantitative measurements of various light field properties on LFSD datasets. (a) PR curves; (b) ROC curves.

Model	F-measure	AUC	MAE
Color	0.5923	0.7089	0.2367
Color+Bg	0.6390	0.7708	0.2157
Depth	0.5587	0.7354	0.2421
Depth+Bg	<u>0.7297</u>	<u>0.8676</u>	<u>0.1708</u>
Color+Depth	0.6422	0.7904	0.2255
Color+Depth+Bg	**0.7749**	**0.8982**	**0.1605**

Table 1: Comparisons of F-measure, ROC and MAE from different light field properties (bold: best; underline: second best).

saliency, our approach outperforms the versions with individual ones, suggesting its ability to leverage diversified light field cues. Furthermore, the performance is significantly improved by computing the background probability from the focusness which is additional light field support of our approach.

Table 1 shows F-measure, ROC and MAE results for comparisons. There is a consistent improvement in performance for all the metrics, which further validates the effectiveness of our approach for light field saliency detection.

Figure 4 visually compares the performance of different light field properties. We observe that each of cues has its unique advantage to saliency detection in different ways. Depth cue is exploited to detect foreground salient objects. However, it may fail when the depth contrast is low or salient object is placed in the background, *e.g.*, the 2nd example. In this example, the color cue from the all-focus image can be used to distinguish salient and non-salient colors in the entire scene. Further, the focusness at different depth levels is beneficial for efficient foreground and background separation. The 3rd and 4th examples show that background priors encoded by light field focusness are helpful to eliminate the background distraction and enhance salient foreground objects.

Extending 2D Saliency Models with Light Field Depth-induced Saliency

To validate the benefit of light field depth, we extend 8 state-of-the-art 2D saliency approaches by fusing 2D saliency maps with light field depth contrast saliency maps into final ones through the standard pixel-wise summation. These methods include Tavakoli [Tavakoli *et al.*, 2011], CNTX [Goferman *et al.*, 2012], GS [Wei *et al.*, 2012], SF [Perazzi *et al.*, 2012],
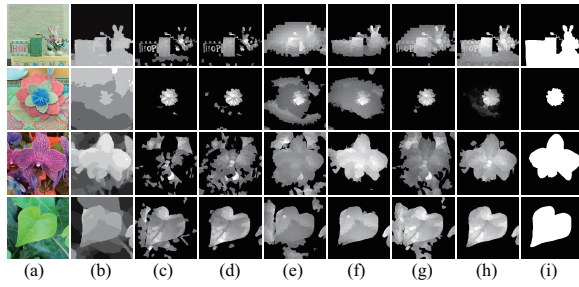
Figure 4: Visual comparisons of saliency estimation from different light field properties. (a) all-focus image; (b) depth map; (c) color; (d) color+bg; (e) depth; (f) depth+bg; (g) color+depth; (h) color+depth+background; (i) GT.
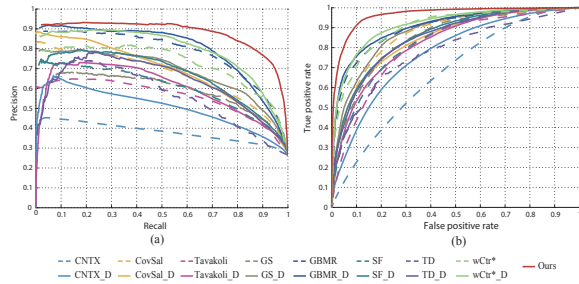

Figure 5: Quantitative results of our approach, state-of-the-art 2D approaches and their depth-extended versions. (a) PR curves; (b) ROC curves.

TD [Scharfenberger *et al.*, 2013], CovSal [Erdem and Erdem, 2013], GBMR [Yang *et al.*, 2013], and wCtr* [Zhu *et al.*, 2014]. We set the models all with default parameters in their original implementations.

Figure 5 presents the PR and ROC curves of our results. The comparisons of F-measure, ROC and MAE are given in Table 2. Here the postfix '_D' denotes depth-extended saliency methods. We can see that our approach is superior to all the state-of-the-art 2D models even combined with the light field depth-induced saliency. The accuracy from all the 2D saliency methods are improved by incorporating the light field depth saliency by about 1–5% and 3–6% for F-measure and MAE, respectively. It is worth to note that we obtain the significant improvement for CNTX by 10% in the AUC metric.

Figure 6(c)–6(j) show qualitative comparisons of all the 2D saliency methods (Top) and their depth-extended versions (Bottom) for two examples. It is obvious that most approaches fail when the object has the similar appearance as the background or the background is cluttered. However, the inclusion of the light field depth contrast helps to capture homogeneous color elements and subtle textures within the object so as to identify foreground salient objects.

We also visually compare our approach (Figure 7(n)) with all the depth-extended approaches (Figure 7(c)–7(j)). Benefiting from the combination of color and depth contrast with background priors, our approach still efficiently works when

Model	F-measure	AUC	MAE
CNTX	0.3643	0.6700	0.3574
CNTX_D	0.4123	0.7718	0.3514
CovSal	0.6335	0.8599	0.2417
CovSal_D	0.6373	0.8466	0.2850
Tavakoli	0.5498	0.8078	0.2551
Tavakoli_D	0.5711	0.8276	0.2903
GS	0.5944	0.8443	0.2395
GS_D	0.6217	0.8792	0.2843
GBMR	0.7461	0.8965	<u>0.1822</u>
GBMR_D	0.7536	0.9072	0.2415
SF	0.4678	0.8301	0.2468
SF_D	0.4704	0.8552	0.2903
TD	0.5766	0.7775	0.2623
TD_D	0.5999	0.8490	0.2951
wCtr*	0.6996	0.8991	0.1878
wCtr*_D	0.7382	0.9156	0.2475
DVS	0.2723	0.5354	0.3274
DVS_Bg	0.2851	0.5509	0.2846
ACSD	0.7905	<u>0.9467</u>	0.1830
ACSD_Bg	<u>0.8025</u>	0.8361	0.1668
LFS	0.7500	0.9272	0.2077
Ours	**0.8186**	**0.9641**	**0.1363**

Table 2: Comparisons of F-measure, AUC, and MAE from our approach, state-of-the-art 2D/3D approaches and their light field-extended methods (bold: best; underline: second best).
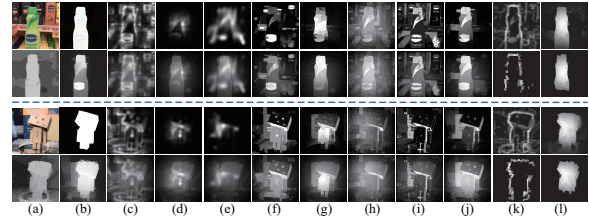

Figure 6: Visual comparisons of 10 state-of-the-art 2D/3D saliency detection models and their light field-extended versions for two examples. (a) all-focus image (Top) and depth map (Bottom); (b) GT (Top) and ours (Bottom); (c) CNTX; (d) CovSal; (e) Tavakoli; (f) GS; (g) GBMR; (h)SF; (i)TD; (j) wCtr*; (k) DVS; (l) ACSD.

the background is not distant enough or the salient object is not distinct at the depth level.

Extending 3D Saliency Models with Light Field Focusness-induced Background Priors

In order to show the role of light field focusness, we incorporate background priors (Eq. 5) computed from focusness maps into 2 state-of-the-art 3D saliency models: DVS [Ciptadi *et al.*, 2013] and ACSD [Ju *et al.*, 2014].

Similarly, the quantitative results are shown in Figure 8 and Table 2. The postfix '_Bg' indicates the methods extended with background priors from the light field focusness. Overall, the background prior encoded by the focusness cue improves original 3D saliency detection, which can also be seen in Figure 6(k) and (l) for visual comparisons. Apparently,
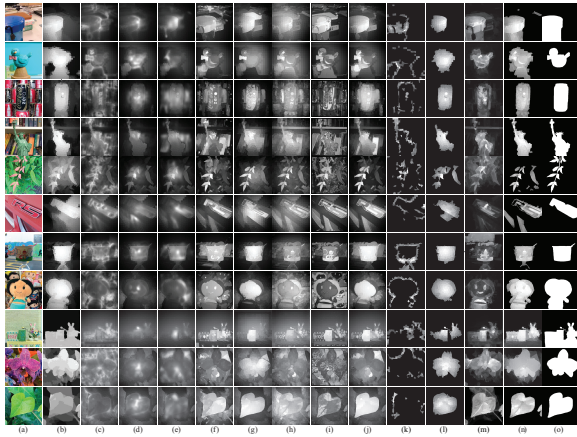
Figure 7: Visual comparisons of our approach and 2D/3D extended methods. (a) all-focus image; (b) depth map; (c) CNTX_D; (d) CovSal_D; (e) Tavakoli_D; (f) GS_D; (g) GBMR_D; (h)SF_D; (i) TD_D; (j) wCtr*_D; (k) DVS_Bg; (l) ACSD_Bg; (m) LFS; (n) Ours; (o) GT.
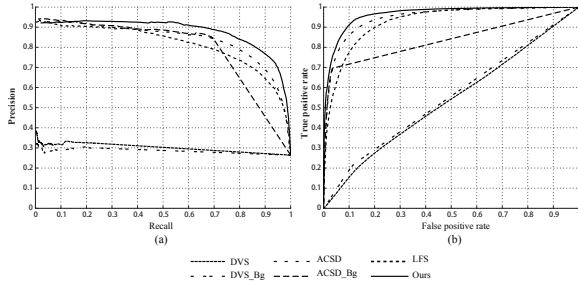

Figure 8: Quantitative results of our approach, LFS, state-of-the-art 3D approaches and their focusness-extended versions. (a) PR curves; (b) ROC curves.

DVS pays more attention on object contours while ignores inner salient regions. ACSD is the second best approach in our comparisons because they also consider the depth contrast and location cue which is very beneficial for saliency detection. Additionally, Figure 7 visually shows that our approach performs better than these extended approaches (Figure 7(k) and (l)).

Comparison with the LFS Method

LFS [Li *et al.*, 2014] is so far the only reported approach for saliency detection on the light field data. As can be seen from Figure 8, our approach achieves much better performance than LFS at the higher precision rates in PR curves and the lower false positive rates in ROC curves. Figure 7(m) and (n) show the qualitative comparisons of LFS and our approach. Our method locates foreground objects more accurately. This is mainly because the depth contrast saliency provides us strong salient cue in cluttered background or the background having the similar colors with the foreground. Besides, background regions are falsely detected as salient objects by LFS in some cases while ours not (see the 5th, 7th and 11th examples). The possible reason could be that our ap-

proach makes an appropriate use of focal slices to introduce the background probability on the focusness map, which is beneficial for saliency detection when salient objects cannot be distinguished at the depth level.

5 Conclusions

The light field camera allows one to capture the total amount of light intensity and the direction of each ray from incoming light in a single exposure simultaneously, which provides not only intensity information, but also depth and focusness information. These interesting properties of the light field motivated us to investigate its capabilities for visual saliency. In this paper, we proposed a new saliency detection approach using light field focusness, depth and all-focus cues. Our approach produced state-of-the-art saliency maps on the LFSD dataset. Through extensive evaluations, we showed that various 2D approaches supported by our light field depth-induced saliency improved their accuracy of saliency detection, and by considering different focus areas from the light field, backgrounds are easily separated from foregrounds, which is beneficial for 3D saliency detection. Compared with [Li *et al.*, 2014], our approach achieved substantial gains in accuracy. However, the depth reliefs of the light field are limited in the current cameras. In future work, we are interested in making use of light field depth and focusness to predict gaze shifts in real 3D scenes and estimating more accurate depth maps from light field cameras with large depth reliefs to improve visual saliency applications.

References

[Achanta *et al.*, 2009] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk. Frequency-tuned salient region detection. In *CVPR*, pages 1597–1604, 2009.

[Achanta *et al.*, 2012] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *TPAMI*, 34(11):2274–2282, 2012.

[Adelson and Wang, 1992] Edward Adelson and John Wang. Single lens stereo with a plenoptic camera. *TPAMI*, 14(2):99–106, 1992.

[Borji *et al.*, 2012] Ali Borji, Dicky N. Sihite, and Laurent Itti. Salient object detection: A benchmark. In *ECCV*, pages 414–429, 2012.

[Cheng *et al.*, 2011] Ming-Ming Cheng, Guo-Xin Zhang, Niloy Mitra, Xiaolei Huang, and Shi-Min Hu. Global contrast based salient region detection. In *CVPR*, pages 409–416, 2011.

[Cheng *et al.*, 2013] Ming-Ming Cheng, Jonathan Warrell, Wen-Yan Lin, Shuai Zheng, Vibhav Vineet, and Nigel Crook. Efficient salient region detection with soft image abstraction. In *ICCV*, pages 1529–1536, 2013.

[Ciptadi *et al.*, 2013] Arridhana Ciptadi, Tucker Hermans, and James M Rehg. An in depth view of saliency. In *BMVC*, pages 9–13, 2013.

[Desingh *et al.*, 2013] Karthik Desingh, K Madhava Krishna, Deepu Rajan, and CV Jawahar. Depth really matters: Improving visual salient region detection with depth. In *BMVC*, pages 98.1–98.11, 2013.

[Erdem and Erdem, 2013] Erkut Erdem and Aykut Erdem. Visual saliency estimation by nonlinearly integrating features using region covariances. *Journal of vision*, 13(4):11, 2013.

[Goferman *et al.*, 2012] Stas Goferman, Lihi Zelnik-Manor, and Ayellet Tal. Context-aware saliency detection. *TPAMI*, 34(10):1915–1926, 2012.

[Gortler *et al.*, 1996] Steven J Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F Cohen. The lumigraph. In *SIGGRAPH*, pages 43–54, 1996.

[Han and Vasconcelos, 2014] Sunhyoung Han and Nuno Vasconcelos. Object recognition with hierarchical discriminant saliency networks. *Front. Comput. Neurosci.*, 8(doi: 10.3389/fncom.2014.00109):109, 2014.

[Itti and Koch, 2001] Laurent Itti and Christof Koch. Computational modelling of visual attention. *Nature reviews neuroscience*, 2(3):194–203, 2001.

[Itti *et al.*, 1998] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *TPAMI*, 20(11):1254–1259, 1998.

[Jansen *et al.*, 2009] Lina Jansen, Selim Onat, and Peter König. Influence of disparity on fixation and saccades in free viewing of natural scenes. *Journal of Vision*, 9(1):1–19, 2009.

[Jia and Han, 2013] Yangqing Jia and Mei Han. Category-independent object-level saliency detection. In *ICCV*, pages 1761–1768, 2013.

[Jiang *et al.*, 2013] Peng Jiang, Haibin Ling, Jingyi Yu, and Jingliang Peng. Salient region detection by ufo: Uniqueness, focusness and objectness. In *ICCV*, pages 1976–1983, 2013.

[Ju *et al.*, 2014] Ran Ju, Ling Ge, Wenjing Geng, Tongwei Ren, and Gangshan Wu. Depth saliency based on anisotropic center-surround difference. In *ICPR*, pages 1115–1119, 2014.

[Koch and Ullman, 1985] Christof Koch and Shimon Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Human neurobiology*, 4(4):217–229, 1985.

[Lang *et al.*, 2012] Congyan Lang, Tam V Nguyen, Harish Katti, Karthik Yadati, Mohan Kankanhalli, and Shuicheng Yan. Depth matters: Influence of depth cues on visual saliency. In *ECCV*, pages 101–115, 2012.

[Li *et al.*, 2013] Xiaohui Li, Huchuan Lu, Lihe Zhang, Xiang Ruan, and Ming-Hsuan Yang. Saliency detection via dense and sparse reconstruction. In *ICCV*, pages 2976–2983, 2013.

[Li *et al.*, 2014] Nianyi Li, Jinwei Ye, Yu Ji, Haibin Ling, and Jingyi Yu. Saliency detection on light field. In *CVPR*, pages 2806–2813, 2014.

[Li, 2002] Zhaoping Li. A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, 6(1):9–16, 2002.

[Lumsdaine and Georgiev, 2009] Andrew Lumsdaine and Todor Georgiev. The focused plenoptic camera. In *ICCP*, pages 1–8, 2009.

[Mahadevan and Vasconcelos, 2012] Vijay Mahadevan and Nuno Vasconcelos. On the connections between saliency and tracking. In *NIPS*, pages 1664–1672, 2012.

[Nagahara *et al.*, 2008] Hajime Nagahara, Sujit Kuthirummal, Changyin Zhou, and Shree K Nayar. Flexible depth of field photography. In *ECCV*, pages 60–73, 2008.

[Ng *et al.*, 2005] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan. Light field photography with a hand-held plenoptic camera. *CSTR*, 2(11), 2005.

[Peng *et al.*, 2014] Houwen Peng, Bing Li, Weihua Xiong, Weiming Hu, and Rongrong Ji. Rgbd salient object detection: A benchmark and algorithms. In *ECCV*, pages 92–109, 2014.

[Perazzi *et al.*, 2012] Federico Perazzi, Philipp Krahenbuhl, Yael Pritch, and Alexander Hornung. Saliency filters: Contrast based filtering for salient region detection. In *CVPR*, pages 733–740, 2012.

[Reynolds and Desimone, 2003] John H Reynolds and Robert Desimone. Interacting roles of attention and visual salience in V4. *Neuron*, 37(5):853–863, 2003.

[Rutishauser *et al.*, 2004] Ueli Rutishauser, Dirk Walther, Christof Koch, and Pietro Perona. Is bottom-up attention useful for object recognition? In *CVPR*, pages 37–44, 2004.

[Scharfenberger *et al.*, 2013] Christian Scharfenberger, Alexander Wong, Khalil Fergani, John S Zelek, and David A Clausi. Statistical textural distinctiveness for salient region detection in natural images. In *CVPR*, pages 979–986, 2013.

[Shi *et al.*, 2014] Jianping Shi, Li Xu, and Jiaya Jia. Discriminative blur detection features. In *CVPR*, pages 2965–2972. IEEE, 2014.

[Tao *et al.*, 2013] Michael W Tao, Sunil Hadap, Jitendra Malik, and Ravi Ramamoorthi. Depth from combining defocus and correspondence using light-field cameras. In *ICCV*, pages 673–680, 2013.

[Tavakoli *et al.*, 2011] Hamed Rezazadegan Tavakoli, Esa Rahtu, and Janne Heikkilä. Fast and efficient saliency detection using sparse sampling and kernel density estimation. In *SCIA*, pages 666–675, 2011.

[Wei *et al.*, 2012] Yichen Wei, Fang Wen, Wangjiang Zhu, and Jian Sun. Geodesic saliency using background priors. In *ECCV*, pages 29–42, 2012.

[Yang *et al.*, 2013] Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang. Saliency detection via graph-based manifold ranking. In *CVPR*, pages 3166–3173, 2013.

[Zhang *et al.*, 2010] Yun Zhang, Gangyi Jiang, Mei Yu, and Ken Chen. Stereoscopic visual attention model for 3d video. In *MMM*, pages 314–324, 2010.

[Zhang *et al.*, 2014] Rumin Zhang, Yu Ruan, Dijun Liu, and Youguang Zhang. All-focused light field image rendering. In *Pattern Recognition*, volume 484 of *Communications in Computer and Information Science*, pages 32–43, 2014.

[Zhang, 2012] Zhengyou Zhang. Microsoft kinect sensor and its effect. *IEEE MultiMedia*, 19(2):4–10, April 2012.

[Zhu *et al.*, 2014] Wangjiang Zhu, Shuang Liang, Yichen Wei, and Jian Sun. Saliency optimization from robust background detection. In *CVPR*, pages 2814–2821, 2014.