Saliency Detection on Light Field report

Paper's Authors

Nianyi Li, Jinwei Ye, Yu Ji, Haibin Ling, and Jingyi Yu

Advisor

Dr. Maryam Abedi

Student

Mohammad Shahpouri

November
2022

Contents

List of Figures

List of Equations

1 Computing Light Field Saliency Cues

Saliency detection approach using the light field is shown in Figure 1.

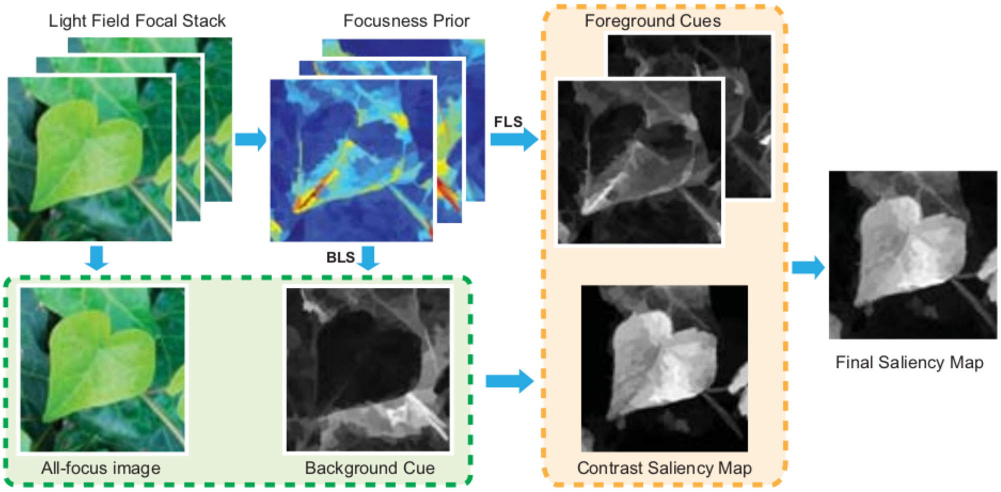

Figure 1. Processing pipeline of our saliency detection algorithm for light fields.

1.1 Focal Stack and All-Focus Images

A Lytro light field camera is used to synthesize a focal stack and further a all-focus image. All-focus image is composed by focus fusion using existing online-tools[1] from the focal stack.

Notaions are: $\{I^i\}, i = 1, \ldots, N$ the focal stack synthesized from the light field and I^* the all-focus image by fusing the focused of $\{I^i\}$. The goal is to compute a saliency map w.r.t. I^*. Using the mean-shift algorithm [1] to segment each slice $\{I^i\}$ and $\{I^*\}$. Using (x, y) to index a pixel and r to index to a region.

1.2 Focusness Measure

To detect the in-focus regions an $n \times n$ image I is transformed into frequency domain by the Discrete Cosine Transform (DCT):

$$\mathcal{D}(u, v) = \sum_{x=0}^{n-1} \sum_{y=0}^{n-1} \cos(\frac{\pi u}{2n}(2x+1)) \cos(\frac{\pi v}{2n}(2y+1)) I(x, y) \tag{1}$$

a series of M bandpass filters $\{P_m\}, m = 1, \ldots, M$ are applied on $\mathcal{D}(u, v)$ for decomposing the signal and then transform the decomposed results back via the inverse DCT. A sliding window of 8×8 pixels is used and compute

[1] http://code.behnam.es/python-lfp-reader/

the variance τ_m within each patch with respect to filter P_m. To ensure reliable focusness measurements, the harmonic variance [2] is utilized to measure the overall variance over all M filters:

$$\mathcal{F}(x,y) = \left[\frac{1}{M-1} \sum_{m=1}^{M} \frac{1}{\tau_m^2(x,y)} \right]^{-1} \tag{2}$$

$\mathcal{F}(x,y)$ shows the focusness measure at pixel (x,y). To measure the focusness of a region, the average of all pixels within a region r is computed:

$$\mathcal{F}(r) = \sum_{(x,y)\in r} \frac{\mathcal{F}(x,y)}{A_r} \tag{3}$$

where A_r is the total number of pixels in r.

1.3 Background Selection

The focusness measure $\mathcal{F}$ of all pixels along the x and y axes respectively is integrated to form two 1D focusness distributions to select the background:

$$D_x = \tfrac{1}{\alpha} \textstyle\sum_{y=1}^{h} \mathcal{F}(x,y), \quad D_y = \tfrac{1}{\alpha} \textstyle\sum_{x=1}^{w} \mathcal{F}(x,y) \tag{4}$$

where w and h are the width and height of the image and $\alpha = \sum_x \sum_y \mathcal{F}(x,y)$ is the normalization factor.

To quantitatively measure if a focal slice corresponds to the background a "U-shaped" 1D band suppression filter is defined:

$$\mathcal{U}(x,w) = \left(\frac{1}{\sqrt{1+(x/\eta)^2}} + \frac{1}{\sqrt{1+((w-x)/\eta)^2}}\right) \tag{5}$$

where η controls the suppression bandwidth in $\mathcal{U}$ depending on the image size/resolution. Since The Lytro focal stack images have a uniform resolution of 360×360, $\eta = 28$ in all experiments.

To compute a Background Likelihood Score (BLS) for each focal slice I^i the focusness distribution is scaled by the suppression filter:

$$BLS(I^i) = \rho \cdot \left[\sum_{x=1}^{w} D_x^i(x) \cdot \mathcal{U}(x,w) + \sum_{y=1}^{h} D_y^i(y) \cdot \mathcal{U}(y,h)\right] \tag{6}$$

where $\rho = \exp(\frac{\lambda \cdot i}{N})$ is the weighting factor of layer i in terms of depth, N is the total number of slices in the focus stack and $\lambda = 0.2$. The slice with the highest BLS is chosen as the background slice I^B.

1.4 Objectness and Foreground Measures

To measure the objectness of a focused region in a focal stack image I^i, using a 1D gaussian filter with mean μ and variance σ as:

$$\mathcal{G}(x) = \exp(-\frac{x-\mu}{2\sigma^2}) \tag{7}$$

$\mu = x_p$ or y_p corresponds to the centroid of the object and $\sigma = 45$ as its size.

The objectness score (OS) for each focal slice is computed as:

$$OS(I^i) = \sum_{x=1}^{w} D_x^i \cdot \mathcal{G}(x,w) + \sum_{y=1}^{h} D_y^i \cdot \mathcal{G}(y,h) \tag{8}$$

If a region consists of a salient object, it should have a low BLS and high OS, indicating it belongs to the foreground. Therefore a foreground likelihood score (FLS) defines as:

$$FLS(I^i) = OS(I^i) \cdot (1 - BLS(I^i)) \tag{9}$$

The foreground slices $\{I^F\}$ as one with the higher FLS.

2 Saliency Detection

The cues obtained from the light field focal stack are combined to detect saliency in the all-focus image $\{I^*\}$.

Location Cues. First, locating the background regions in $\{I^*\}$. To incorporate the location prior [3], the focusness measure for each region R_r is scaled in terms of its distance to the center of the image and use it as a new background cue:

$$BC(r) = \frac{1}{\gamma}[\mathcal{F}^B(r)] \cdot \|\mathbf{P}_r - \mathbf{c}\|^2 \tag{10}$$

where γ is a normalization factor, $\mathbf{P}_r$ is the centroid of r and $\mathbf{c}$ is image center. BC is thresholded to determine the background regions $\{B_{r'}\}$, $r' = 1, \ldots, K$ in I^* (where K is the total number of background regions). Then Location cue computed as:

$$LC(r) = \exp(-\beta \cdot BC(r)) \tag{11}$$

where $\beta = 10$.

Contrast Cues. The color difference $\delta(r, r') = \max\{|red(r) - red(r')|, |green(r) - green(r')|, |blue(r) -$

$blue(r')|\}$ is calculated For each non-background region r and background region r' in I^*. To improve robustness, the harmonic variance of all $\delta(r, r')$ for r is computed:

$$HV(r) = \left[\frac{1}{K} \sum_{r'=1}^{K} \frac{1}{\delta(r, r')} \right]^{-1} \tag{12}$$

Combining the harmonic variance of color difference HV with location cue LC obtains a color contrast based saliency map as:

$$S_C(r) = HV(r) \cdot LC(r) \tag{13}$$

Foreground Cues. Combining the focusness maps $\mathcal{F}_j^F(r)$ and the location cue LC to compute foreground cues:

$$S_F^j(r) = \mathcal{F}_j^F \cdot LC(r) \tag{14}$$

Combine. Finally, The objectness measure is used as weight for combining the contrast based saliency map $S_C(r)$ and foreground maps $S_F^j(r)$ as:

$$S(r) = \sum_{j=1}^{L} \omega_j \cdot S_F^j(r) + \omega_C \cdot S_C(r) \tag{15}$$

where $\{\omega_j\}$ and ω_C are the objectness weights calculated by Equation 8.

3 Experimental Results

A dataset of 100 light fields using the Lytro light field camera is collected, 60 indoor scenes and 40 outdoor scenes, to evaluate the performance of the method. They compare their approach with spectral residual (SR [4]), spatiotemporal-cues (LC [5]), graph-based saliency (GB [6]), frequency-tuning (FT [7]), global-contrast (HC and RC [8]), Low Rank Matrix Recovery (LRMR [3]), Graph-Based Manifold Ranking (GBMR [9]), focusness-based (UFO [10]) and Hierarchical Saliency (HS [11]). All these methods have open source code and they use the default parameter.

3.1 Quantitative Result

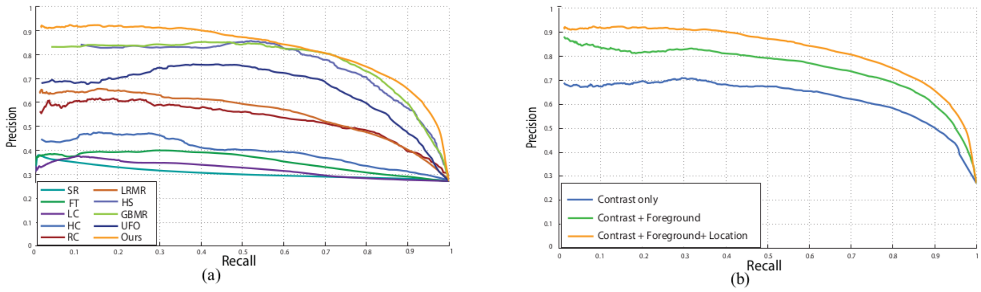

Figure 2. (a) PRC comparisons on the light field dataset; (b) PRC comparisons using different cues in the approach.

3.2 Qualitative Result

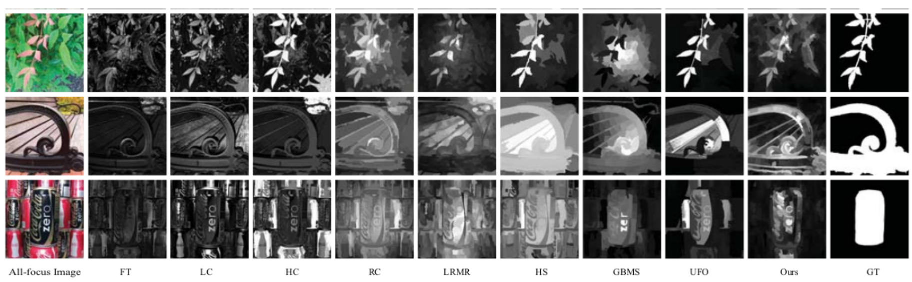

Figure 3. Visual Comparisons of different saliency detection algorithms.

References

[1] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002. 1

[2] F. Li and F. Porikli, "Harmonic variance: A novel measure for in-focus segmentation," in *Proceedings of the British Machine Vision Conference*, pp. 33.1–33.11, BMVA Press, 2013. 2

[3] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 853–860, 2012. 3, 4

[4] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2007. 4

[5] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proceedings of the 14th ACM International Conference on Multimedia*, MM '06, (New York, NY, USA), p. 815–824, Association for Computing Machinery, 2006. 4

[6] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in Neural Information Processing Systems* (B. Schölkopf, J. Platt, and T. Hoffman, eds.), vol. 19, MIT Press, 2006. 4

[7] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1597–1604, 2009. 4

[8] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *CVPR 2011*, pp. 409–416, 2011. 4

[9] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3166–3173, 2013. 4

[10] P. Jiang, H. Ling, J. Yu, and J. Peng, "Salient region detection by ufo: Uniqueness, focusness and objectness," in *2013 IEEE International Conference on Computer Vision*, pp. 1976–1983, 2013. 4

[11] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1155–1162, 2013. 4