Salience Guided Depth Calibration for Perceptually Optimized Compressive Light Field 3D Display

Shizheng Wang[1], Wenjuan Liao[1], Phil Surman[2], Zhigang Tu[2], Yuanjin Zheng[2], and Junsong Yuan[3]

[1]Institute of Microelectronics of CAS
[2]Nanyang Technological University
[3]CSE department, University at Buffalo

jsyuan@buffalo.edu

Abstract

Multi-layer light field displays are a type of computational three-dimensional (3D) display which has recently gained increasing interest for its holographic-like effect and natural compatibility with 2D displays. However, the major shortcoming, depth limitation, still cannot be overcome in the traditional light field modeling and reconstruction based on multi-layer liquid crystal displays (LCDs). Considering this disadvantage, our paper incorporates a salience guided depth optimization over a limited display range to calibrate the displayed depth and present the maximum area of salience region for multi-layer light field display. Different from previously reported cascaded light field displays that use the fixed initialization plane as the depth center of display content, our method automatically calibrates the depth initialization based on the salience results derived from the proposed contrast-enhanced salience detection method. Experiments demonstrate that the proposed method provides a promising advantage in visual perception for the compressive light field displays from both software simulation and prototype demonstration.

1. Introduction

Technologies based on 3D have been increasingly applied in commercial products. However, these technologies, based on binocular parallax using lenticular sheets or parallax barriers suffer from discrepancy between visual accommodation and convergence; this can cause visual confusion and fatigue as the eyes focus on the screen but converge at the apparent distance of the image. To solve this problem, volumetric displays [38, 46] and holographic displays [32] have been developed. A huge amount of data would be involved to provide a true 3D image with both correct fo-
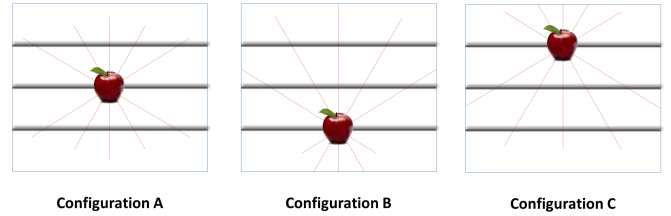

Figure 1. Different depth initialization for the same target light field, here the configuration A is used by [45, 18], configuration B is from [13] and configuration C is [28].

cus and parallax cues and this prevents these technologies from being widely accepted for daily usage. In addition to these pure optical or physics-based techniques, a computational 3D display technology known as compressive light field display has also been investigated [15, 45, 14] for solving the problem. They not only take the advantages of various emerging display technologies such as multi-layer displays and high-speed temporal modulation, but also make use of promising computer vision and pattern recognition algorithms such as sparse reconstruction and compressive factorization which effectively compress the display data from multi-view images to multi-layer images. Moreover, the compressive light field 3D display is compatible with 2D displays since parallax barriers and lenticular sheets are not used in this display method.

However, this kind of computational multi-layer display, including our three-layer light field 3D display, shares the same limitation; the maximum depth range of multi-layer display theoretically has an upper bound and cannot be more than twice the distance between the outer layers [23]. Considering the unsatisfactory depth range limitation of multi-layer liquid crystal displays (LCDs) and the bigger depth range in the light field capture, some depth

range may have to be ignored in the light field modeling and reconstruction, this paper has proposed a salience-guided depth calibration for perceptually optimized compressive light field 3D display.

In this paper, we use a three-layer prototype as an example, but the proposed method also can be extended to other multi-layer displays with more LCD layers. In contrast to existing work on cascaded light field displays that only use the fixed depth initialization configuration [45, 13, 28] as shown in Fig. 1, the proposed method will firstly detect the higher salience objects in the light field capture region based on a proposed contrast enhanced salience detection method, then maps these objects into the compressive display depth region of multi-layer LCDs as much as possible. The framework of the proposed salience guided depth calibration for optimized display is shown in Fig. 2. The experimental results show that an improved display is produced by the proposed optimization framework.

In this paper, Section 2 introduces the related work of salience detection and compressive display of light fields and Section 3 discusses the contrast enhanced salience detection method. Section 4 proposes salience-guided depth calibration for three-layer light field displays. The experimental results are reported in Section 5 and Section 6 summarizes the existing contributions and anticipates future work.

2. Related Works

Before proposing the optimized framework and methods, related work regarding salience detection [6] and multi-layer light field display [44] is introduced in this section.

2.1. Salience Detection on Light Field

With vs. Without Learning. Recently, several studies have developed learning methods in saliency analysis [50, 27, 21]. For example, Judd et al. [21] use the standard support vector machine (SVM) classifier to formulate saliency detection as a binary classification problem. Li et al. [27] propose a saliency detection framework using dense and sparse coding representations as features and integrate this framework via the Bayes formula. Multilayer learning networks [50, 35] approaches also have been researched to acquire the hierarchies of representations to detect the salience objects. However, these approaches suffer from parameter sensitivity and normally require large-scale manually annotated data to train effective models.

State-of-the-art salience detection approaches for light field with the exception of Li et al. [25], use the image's selected foreground to build the dictionaries, and detect the salience objects with a weighted sparse coding framework. Most other works still use the scheme without machine learning. A combined light field salience detection algorithm based on foreground, background and contrast cues is proposed in [26]. Zheng et al. [49] further propose a deeper light field salience detection method for light field by inducing the depth image into the optimization framework. To avoid the parameter sensitivity and manual annotation, this paper implements light field salience detection by a framework without learning.

Foreground vs. Background Priors. Many saliency detection schemes exploit contrast cues, i.e., salience objects are expected to exhibit high foreground contrast within certain context [39]. Koch and Itti [19] use center-surround foreground contrast of low level features to detect saliency. More local methods to compute the foreground contrast within a small neighborhood of pixels are proposed by using color difference [7], edge orientations [29], or curvatures [40]. Global methods, considering statistics of the entire image and rely on features such as power spectrum [16], color histogram [9] and element distributions [36], are also proposed for foreground contrast detection.

Although the center-surround foreground approaches are proven highly effective, Wei et al. [42] suggest that background priors are equally important. In fact, one can eliminate the background to significantly improve foreground detection. Yang et al. [48] observe that connectivity is an important characteristic of background and use a graph-based ranking scheme. Since most existing approaches rely on color contrast, when the foreground and background have similar color, these approaches can easily fail. Thus, our approach resolves this issue by using a contrast enhanced salience detection method on the light field based on selected color contrast cues, depth contrast cues and background cues.

Intrinsic vs. Extrinsic cues. When detecting salience objects, a key step is to distinguish salience targets from distractors. Toward this end, some approaches propose to extract various cues only from the input image itself to pop-out targets and suppress distractors (i.e., the intrinsic cues) [17, 41, 34, 40]. However, other approaches argue that targets and distractors may share some common visual attributes and the intrinsic cues are often insufficient to distinguish them. Therefore, they incorporate extrinsic cues such as user annotations [6], depth map [49] or statistical information [24] of similar images to facilitate detecting salient objects in the image.

This paper will focus on salience detection on light field, in which some important extrinsic cues already appear within the all-focus image, such as depth image, re-focus images and dense multi-view images. Thus the proposed approach could be listed within the same classification as extrinsic cues.
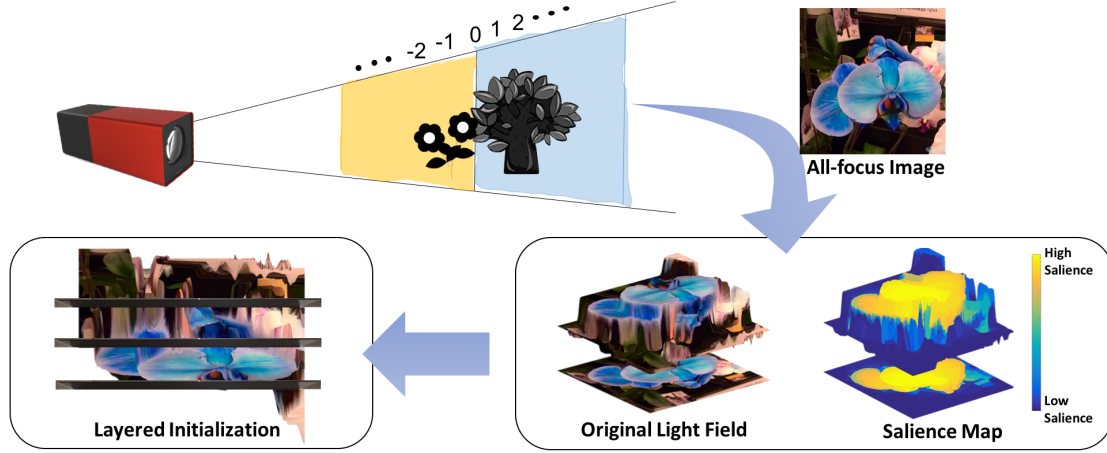
Figure 2. The framework of the salience guided depth calibration for perceptually optimized compressive light field 3D display, including light field capture, saliency detection, depth initialization, layered decompression and perceptually optimized light field display.

2.2. Multi-layer Display

Multi-layer display is a type of extension of normal liquid crystal display which stacks multiple liquid crystal (LC) layers with a uniform backlight as shown in Fig. 3. The traditional additive volume method will show a 3D scene at different depths to generate 3D display [5]. Further, Bell et al. [4] proposed a two-layer 3D display system and discussed the implementation details like moiré reduction by diffuser for this system. For light field display, a theoretical simulation for multi-layer light field display is proposed in [12], and its implementation is introduced in [13]. Compressive display [44] is a kind of multi-layer light field display which has complete theoretical modeling and analysis, as well as several prototype implementations, such as polarization display [22] and tensor display [45].

In this paper, we will follow the light field modeling in [45] to introduce the proposed perceptually optimized compressive light field 3D display. The detailed modeling will be introduced in the following subsection.

2.3. Light Field Modeling

The compressive display is used to depict a discrete light field, which allows for optimal decomposition of a light field into light-attenuating layers [43]. The modeling framework for both glasses-free [45] and head-mounted stereoscope [18] could be considered as a fixed stack of N light-attenuating layers illuminated by a uniform backlight.

As shown in Fig. 3, we demonstrate a three-layer attenuation configuration. The reconstructed light field R for this three-layer tensor display can be written as:

$$[\mathbf{L}^{(1)}, \mathbf{L}^{(2)}, \mathbf{L}^{(3)}] = \mathbf{L}^{(1)}(\lambda_a)\mathbf{L}^{(2)}(\lambda_b)\mathbf{L}^{(3)}(\lambda_c), \quad (1)$$

where $\mathbf{L}^{(n)}(\lambda_i) \in [0, 1]$ is the transmittance at the pixel λ_i
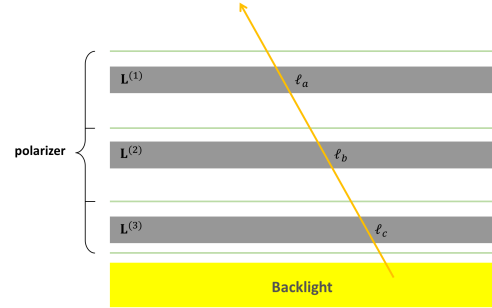

Figure 3. Polarization-based attenuation light field displays.

of layer n. The transmittances for the front, middle and rear layers are given by $\mathbf{L}^{(1)}(\lambda_a)$, $\mathbf{L}^{(2)}(\lambda_b)$ and $\mathbf{L}^{(3)}(\lambda_c)$, respectively, for a light field ray shown in the figure.

By generalizing Eq. (1), the optimal three-layer reconstructed light field can be acquired by solving a constrained optimization problem [22] as follows:

$$\underset{\{0 \leq \mathbf{L}^{(1)}, \mathbf{L}^{(2)}, \mathbf{L}^{(3)} \leq 1\}}{\arg \min} \frac{1}{2} \left\| \mathbf{T} - \mathbf{W} \circ [\mathbf{L}^{(1)}, \mathbf{L}^{(2)}, \mathbf{L}^{(3)}] \right\|^2. \tag{2}$$

Here, $\circ$ is the product of a Hadamard (elementwise) matrix, $\leq$ denotes the element-wise matrix inequality operator and $\mathbf{T}$ denotes the target 3D light field to be reconstructed. $\mathbf{W}$ is a sparse three-order tensor, only with 0/1 elements, which controls the optimized view angle of reconstruction light field [30].

Eq. (2) corresponds to non-negative matrix factorization. Therefore, we can apply any non-negative matrix factorization algorithm to its solution [31, 8]. However, after applying a Fourier transform into this light field expression, Lanman, D., et al. [23] get the theoretical upper-bound

3

depth of field (DoF) of this three-layer display, which is twice separation layer distance. This means that during the light field reconstruction, we have to map the original camera-captured depth range to the limited multi-layer display depth range where the former is much larger than the latter [45, **?**].

3. Contrast Enhanced Salience Detection

The pipeline of the proposed salience detection approach for display optimization is shown in Fig. 2 and the detailed salience detection algorithm is described in the following subsections.

3.1. Preprocessing

The contrast-enhanced saliency is based on the light field color, depth and focus cues on the super-pixel [3]. Here, the super-pixels are segmented from the all-focus image $\widehat{I}$. The minimum preprocessing unit of the following analysis is based on super-pixels. The total number of super-pixels M is set to 300 following [49].

After segmenting the all-focus image to super-pixels, we further try to select a best image with the highest color contrast as well as background cues. Here, we denote $\{I^i\}$, $i = 1, ..., N$ as the re-focus images based on the Lytro camera. Following that, we use (x, y) to index a traditional pixel and p to index a super-pixel region.

We start by detecting the in-focus regions in each re-focus image I^i and use them as the focusness prior. The focus value $\mathcal{F}(x, y)$ is measured at pixel (x, y) based on the focusness detection technique [37]. Thus, the focus value of a super-pixel p can be computed from the average of all pixels within the superpixel p:

$$\mathcal{F}(p) = \sum_{(x,y \in p)} \frac{\mathcal{F}(x, y)}{Z_p}, \qquad (3)$$

where Z_p is the total number of traditional pixels in super-pixel p. Thus, a super-pixel level focusness map, $\mathcal{F}^i$, can be generated for each re-focus image I^i. Furthermore, the focus value of the whole image could be integrated to two 1D focusnesss distributions along the x and y axes as:

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

Next, we set out to find one re-focus image with the highest background cues which could be used to detect the non-salience region as well as show the high contrast with salience region. For this purpose, we compute the background likelihood score $BLS(I^i)$ based on each re-focus image I^i by U-shaped filtering D_x and D_y:

$$BLS(I^i) = \rho \cdot \left[\sum_{x=1}^{w} D_x^i(x) \cdot \mathcal{U}(x, w) + \sum_{y=1}^{h} D_y^i(y) \cdot \mathcal{U}(y, h) \right],$$
$$(5)$$

$$\mathcal{U}(x, w) = \left(\frac{1}{\sqrt{1 + (\frac{x}{\alpha})^2}} + \frac{1}{\sqrt{1 + (\frac{w-x}{\alpha})^2}} \right). \quad (6)$$

Here, this filter has a 1D bandpass filtering function along the axis, and $w = 360$ is equal to the width of test images in the Lytro1 dataset, while h is the length and $\alpha = 28$ controls the bandwidth. $\rho = \exp(\lambda * i/N)$, in which $\lambda = 0.2$ is the weighting factor of a layer in terms of depth, and N is the total number of slices in the focus stacks. Then, I^* is chosen from the focusness maps for its highest BLS. While also let the re-focus image $\mathcal{F}^*$ with the highest BLS among focus stacks take the place of all-focus image used in traditional methods [49, 26] as the color contrast cues in super-pixel level. Benefiting from the focused foreground, it will provide a better performance for the final saliency map than an all-focus image.

3.2. Contrast Enhanced Salient Segmentation

After completing the preprocessing work for super-pixel segmentation that is focusness map generation chosen and color image chosen, we integrate focusness background cues, color contrast and depth contrast for generating the final saliency result.

Background Cues. To enhance the saliency contrast, a background probability Pb on the focusness map $\mathcal{F}^*$ is calculated through:

$$Pb(i) = 1 - \exp(-\frac{A_{val}(p_i)^2}{2} \cdot \|C - A_{pos}(p_i)\|^2), \quad (7)$$

where $A_{val}(p_i)$ is the average value of super-pixel p_i on the focusness map $\mathcal{F}^*$ and $A_{pos}(p_i)$ defines normalized average coordinates of super-pixel p_i. $\|C - A_{pos}(p_i)\|$ defines the L2-norm distance of measuring the spatial information of super-pixels related to the image center C. Therefore, regions that belong to the background have higher background probability Pb on $\mathcal{F}^*$.

Contrast. Here, we use the similar method to calculate the color contrast saliency M_C and depth contrast saliency M_D with different inputs, but same processing unit:

$$M_C(i, j) = \|A_{col}(P_i) - A_{col}(P_j)\|D(i, j), \quad (8)$$

4

$$M_D(i,j) = \|A_{dep}(P_i) - A_{dep}(P_j)\|D(i,j), \quad (9)$$

where $i, j = 1, ..., K$, K is the total number of super-pixels. $A_{col}(p_i)$ is the average color value of super-pixel p_i in LAB color space on a selected color re-focus image I^* and $A_{dep}(p_i)$ is the average depth value of p_i on the depth image.

$D(i,j)$ is the spatial factor for controlling the pair-wise distance of super-pixels and is given by:

$$D(i,j) = \exp(-\frac{\|A_{pos}(p_i) - A_{pos}(p_j)\|^2}{2\sigma_w^2}), \quad (10)$$

where, $A_{pos}(p_i)$ and $A_{pos}(p_j)$ are the normalized average coordinates of super-pixel p_i and p_j, separately. σ_w is specified as 0.67 throughout our experiments.

Combined Saliency Metric. Finally, we incorporate background probability into the contrast enhanced saliency as follows:

$$M = \beta M_D + (1 - \beta)M_C, \quad (11)$$

$$M_{com} = M * P_b. \quad (12)$$

Here, M is from the focussness background cues and S^* is the weighted saliency based on color contrast saliency and depth contrast saliency. $\beta = 0.3$.

Post-optimization. Like the state-of-the-art work [49], we also applied saliency optimization algorithm [51] onto the contrasted enhanced saliency map as post-optimization:

$$S_{opt} = OPT(M_{com}). \quad (13)$$

Here, when use the post-processing, the combined salience map is used as a 2D image input for the post-optimization, and the output is a 2D optimized salience map. Experimental results in Supplementary Materials show that the proposed algorithm can be better than the-state-of-the-art by either using, or not using post-processing.

4. Perceptually Optimized Light Field 3D Display via Salience Guided Depth Initialization

4.1. Salience Guided Depth Initialization

Different to the previous mentioned configurations [45, 13, 28, 18] with manually fixed initialization, the proposed method will automatically determine the relative position of
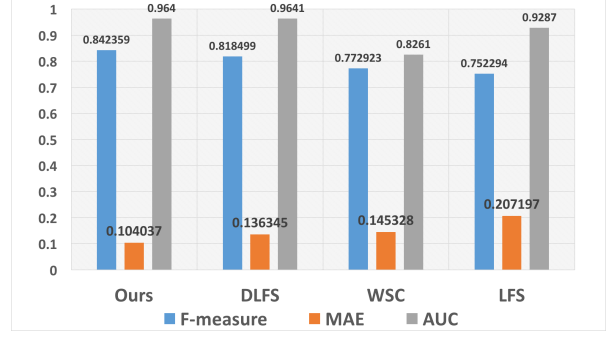

Figure 4. Performance comparisons of ours vs. DLFS, LFS, WSC by F-measure, MAE and AUC.

three layers in order to improve the performance 3D display. The distance between each two layers is fixed to Dis_L at 15 mm.

According to [2], assume the captured depth range is R_C, and the ideal displayed depth range is R_D, thus R_D could be calculated as follows:

$$R_D = \frac{PS_L}{PS_C} \cdot R_C, \quad (14)$$

where the PS_L is the pixel size of the LC screen of our multi-layer light field display prototype, PS_C is the micro-lens size of the Lytro camera related to one light field pixel.

According to [10], assume the real displayed depth range is R_L, and the pixel size of prototype is PS_L, so R_L could be calculated as follows:

$$R_L = \frac{2 \cdot Dis_L}{PS_L}. \quad (15)$$

The performance of 3D display will surely great if $R_L \geq R_D$. However, in general, $R_L < R_D$, therefore, part of the image will be out of range of the real displayed depth (R_L) and resulting in performance decreasing because of the missing key content of the image. That is, the saliency region of the image, which contains the most interested content for human eyes, should be presented as much as possible in the real displayed depth range. The new constrained optimization problem is shown as following:

$$\mathbf{T}_{di} = \arg\max_{T'} S_{dis}(T'). \quad (16)$$

Here, $S_{dis}(T')$ is used to calculate the salience region whose depth range could be initialized in $2 \cdot Dis_L$: that is, substantially between the first layer and the rear layer. $\mathbf{T}_{di}$ is the depth-adjusted target light field. Different to the fixed depth initializations used in [45, 13, 28, 18] which is shown in Fig. 1, separately, $\mathbf{T}_{di}$ is the optimized target light field which will keep the salience objects in the actual displayed depth region as much as possible. For example, if the initialization center of the displayed light field is in the

middle layer, the performance of the proposed display will be equated to three-layer tensor display [45], however the optimization also can handle the cases where the salience objects are not in the center of the DoF of light field as shown in Fig. 2. Here, the Eq. (16) is solved by Simulated Annealing (SA) algorithm [33].

4.2. Light field Reconstruction

After acquiring the perceptually optimized initialization of target light field, we still consider the light field reconstruction problem as a constrained optimization problem for the following reason:

$$\underset{\{0 \leq \mathbf{L}^{(1)}, \mathbf{L}^{(2)}, \mathbf{L}^{(3)} \leq 1\}}{\arg\min} \frac{1}{2} \left\| \mathbf{T}_{di} - \mathbf{W} \circ [\mathbf{L}^{(1)}, \mathbf{L}^{(2)}, \mathbf{L}^{(3)}] \right\|^2,$$

(17)

where, the $\mathbf{T}_{di}$ is the depth-adjusted target light field with the optimized initialization. In this paper, three layered images are generated by least squares with linear constraints and bounds (LSQLIN) [11] and its GPU speed-up version based on the simultaneous algebraic reconstruction technique (SART) algorithm [22] is also released with supporting material.

5. Experimental Results

We compare our approach with state-of-the-art techniques for both salience detection and layered 3D display on a public light field dataset [26].

5.1. Salience Detection

Based on the ground truth (GT) released with the light field dataset [26], we follow the canonical precision-recall curve (PRC), F-measure, mean absolute error (MAE) and area under the curve (AUC) methodologies to evaluate the accuracy of the detected saliency. For details about these evaluation methods we refer the reader to [49]. For objectively showing the benefit of proposed contrast enhanced approach, the parameters setting in our implementation is the same as [49, 26], although the proposed approach is not such sensitive to the fixed parameters. The proposed light field saliency detection results will mainly compare with the algorithms that also use light field as input, which are based on Weighted Sparse Coding (WSC [25]), depth combined contrast (DLFS [49]), and tailored Light Field Salience (LFS [26]), but also has been compared with the algorithms using all-focus image as input, which based on global-contrast (RC [9]), Low Rank Matrix Recovery (LRMR [36]), Graph-Based Manifold Ranking (GBMR [48]), focusness-based (UFO [20]) and Hierarchical Saliency (HS [47]).

Table.1, Fig. 4 and Fig. 6 show the results of the four comparison architectures. The PRC of our unified approach achieves state-of-the-art result, and the best Precision, Recall, F-measure, and MAE results are shown where the proposed approach is used on the public dataset.

Our approach can handle highly challenging cases such as the 'Blue Bird' scene in LFS where the deemed saliency regions have a similar color/texture to the non-saliency regions. Notice that both our precision and recall values are higher than other methods with favorable F-measure and MAE in most cases. This indicates that our algorithm is capable of locating the most salient regions with a high confidence. Fig. 5 shows that our technique also produces more visually available results, e.g., it generates more complete and accurate contours.

5.2. Light Field Display

Hardware Implementation: The hardware prototype was built using three Asus VG248QE 24' LCDs and two NVIDIA Quadro graphic cards, which can demonstrate 1920×1080 resolution images or videos with a 144Hz refresh rate.

The display prototype shown in Fig. 7 uses the three-layer structure of Fig. 3. Following the structure of existing prototypes, the front and middle polarizers are orthogonal, and each polarizer has to be orthogonal to the next. Additionally, the distance between each pair of LCD screens is 15 mm.

Software Simulation: Under the Windows 10 operating system, the software stimulation is implemented based on the light field database [26] captured by a Lytro1 camera, using an Intel CPU Core (3.4 GHz) PC with 32G RAM. Initialized configuration and related parameters are: i) the views, and of each light field content of this dataset is extracted from the raw data (*.LFP) with Lytro desktop software [2] and related toolbox [1], ii) based on the Lytro camera capture configuration, the optimized view angle is [-3, +3] for both horizontal and vertical directions, iii) the time-multiplexed modulation of light field reconstruction is not used in the proposed prototype. Following the above preferences, the average processing time of light field modeling and optimizing steps in MATLAB is approximately 20 minutes.

Fig. **??** shows the average peak signal-to-noise ratio (PSNR) curves of fixed configuration A [45, 18], configuration B [13], and configuration C [28] without the proposed depth initialization optimization (IO) method and the ROC of the proposed IO result respectively. Note that the final simulated result of the proposed IO method is better than any other fixed configuration as it explores all three configurations during the optimization. The layered pictures for each configuration are shown in Fig. 8.
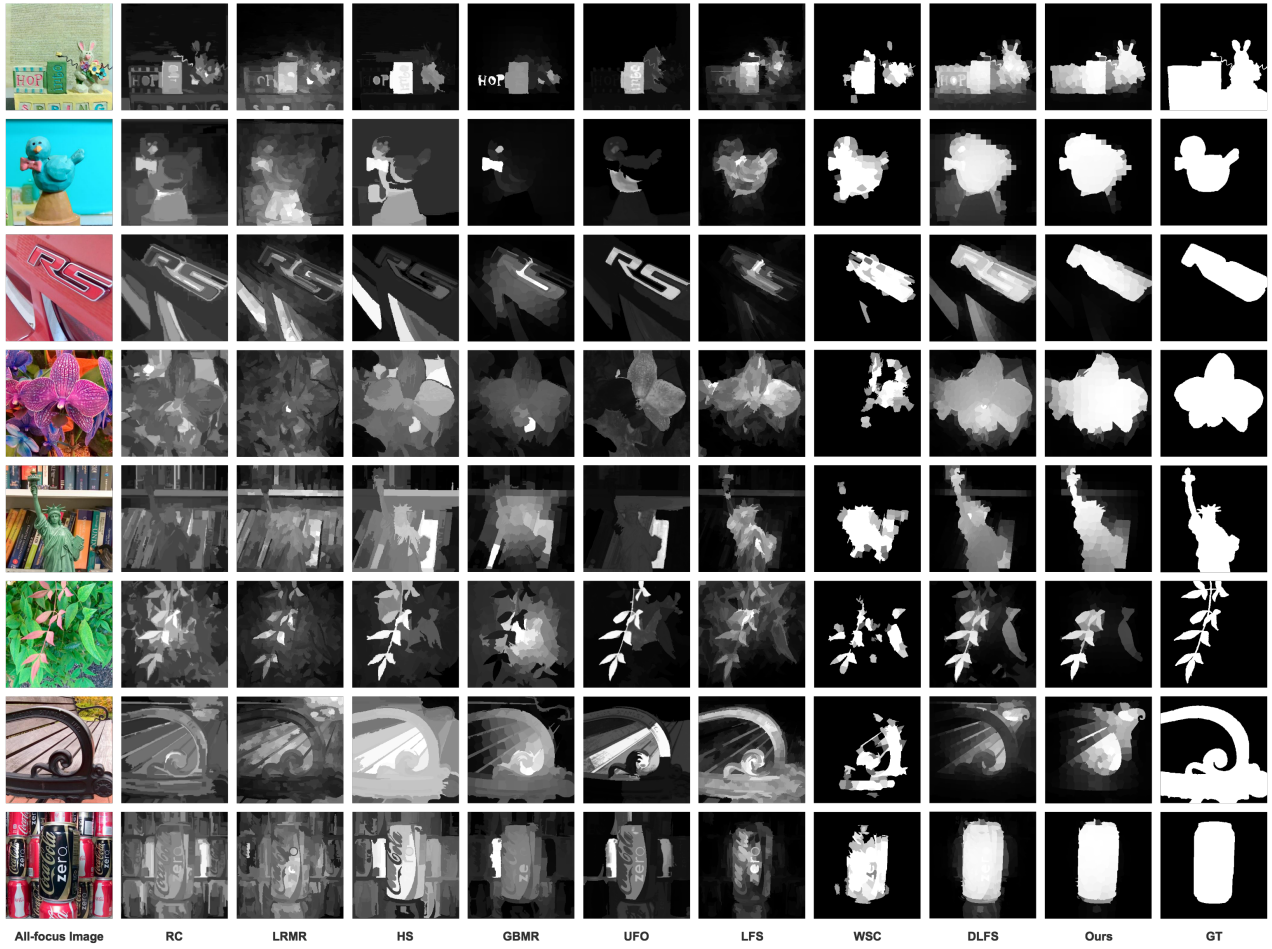
Figure 5. Visual comparisons of different saliency detection algorithms vs. ours on a light field dataset.
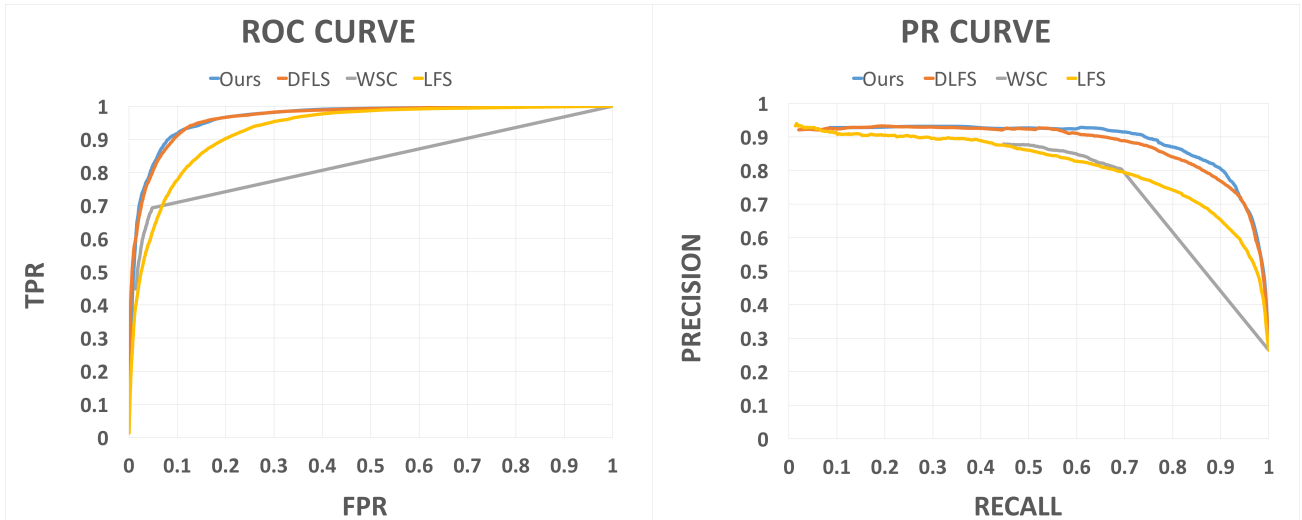

Figure 6. Performance comparisons of ours vs. (a) DLFS, (b) WSC and (c) LFS. The left is the receiver operating characteristic (ROC) curve with true positive rate (TPR) against the false positive rate (FPR) and the right is the precision recall (PR) curve.

	Precision	Recall	F-measure	MAE	AUC
WSC	0.809866	0.670907	0.772923	0.145328	0.8261
LFS	0.810021	0.607887	0.752294	0.207197	0.9287
DLFS	0.827058	0.791206	0.818499	0.136345	**0.9641**
Ours	**0.856693**	0.802748	**0.851966**	**0.097374**	0.897749
Ours+DLFS	0.851779	**0.812412**	0.842359	0.104037	0.9640

Table 1. Evaluation results of ours vs. DLFS, LFS, WSC by Precision, Recall, F-measure, MAE and AUC.

Figure 7. Hardware prototype of compressive light field display.
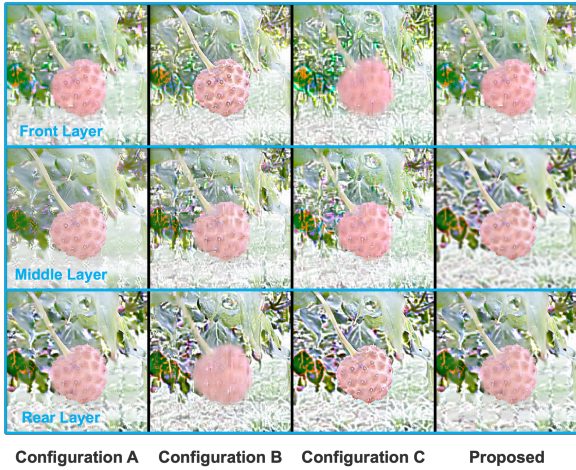

Figure 8. Software-generated layered images for different configurations: the first column is with configuration A, and second column is with configuration B, the third column is based on the configuration C, and the fourth column is from the proposed method. Note that the proposed method's average sharpness is much better than other configurations.

5.3. System Performance

The practical performances of variable prototype configurations with different depth initialization are shown in Fig. 9. Although the visual effect of frontal photographs is limited by luminance and moir, it can be found that our approach provides a more distinct and natural scene than anchor methods.

The proposed method has undergone comprehensive subjective measurement by 12 subjects where the perfor-
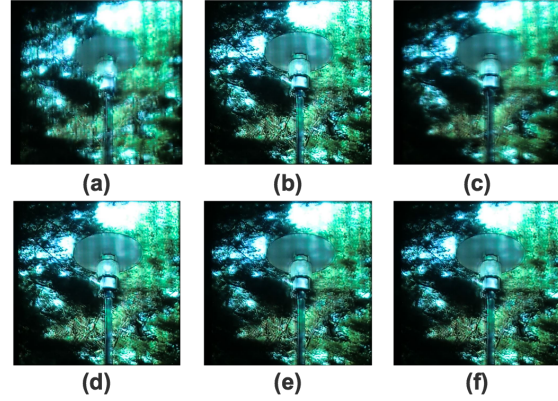

Figure 9. Figure 10. The frontal photograph of overall display performance in different configurations with one content from Lytro1 dataset [26]: here, (a) configuration A [45, 18] (b) configuration B [13], (c) configuration C [28], (d) optimized with proposed salience detection, (e) optimized with saliency ground truth, (f) the original captured image.

mance of optimized depth initialization was viewed as the best with 75% cases, which was found to match people visual perception better, at least 22%, than the other fixed depth initializations, which only match the maximum 53%. More experimental results and supplementary material can be found in the supporting material.

6. Conclusion

This paper has proposed a salience-guided depth optimization for multi-layer light field displays. The best depth initialization is automatically chosen based on the contrast-enhanced salience detection approach and salience-guided depth initialization framework. Experiments demonstrate that the proposed method provides a promising perceptual advantage. In the future, we plan to improve the light field display by introducing a just noticeable differences algorithm to the visual attributes of the optimization framework.

References

[1] Light field toolbox 0.4. http://lightfield-forum.com/tag/lightfield-toolbox/.

[2] Official lytro software. https://www.lytro.com.

[3] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282, 2012.

[4] G. P. Bell, R. Craig, R. Paxton, G. Wong, and D. Galbraith. 25.4: Invited paper: Beyond flat panels multi layer displays with real depth. In *SID Symposium Digest of Technical Papers*, volume 39, pages 352–355. Wiley Online Library, 2008.

[5] S. A. Benton. *Selected papers on three-dimensional displays*, volume 162. Society of Photo Optical, 2001.

[6] A. Borji, M.-M. Cheng, H. Jiang, and J. Li. Salient object detection: A survey. *arXiv preprint arXiv:1411.5878*, 2014.

[7] N. Bruce and J. Tsotsos. Saliency based on information maximization. *Advances in neural information processing systems*, 18:155, 2006.

[8] X. Cao, Z. Geng, T. Li, M. Zhang, and Z. Zhang. Accelerating decomposition of light field video for compressive multilayer display. *Optics express*, 23(26):34007–34022, 2015.

[9] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S.-M. Hu. Global contrast based salient region detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3):569–582, 2015.

[10] O. S. Cossairt, J. Napoli, S. L. Hill, R. K. Dorval, and G. E. Favalora. Occlusion-capable multiview volumetric three-dimensional display. *Applied Optics*, 46(8):1244–1250, 2007.

[11] J. Ding, M. Liu, Q. Zhong, H. Li, and X. Liu. Optimization algorithm of near-eye light field displays based on human visual characteristics. *Chinese Optics Letters*, 14(4):041101, 2016.

[12] H. Gotoda. A multilayer liquid crystal display for autostereoscopic 3d viewing. In *IS&T/SPIE Electronic Imaging*, pages 75240P–75240P. International Society for Optics and Photonics, 2010.

[13] H. Gotoda. Implementation and analysis of an autostereoscopic display using multiple liquid crystal layers. In *IS&T/SPIE Electronic Imaging*, pages 82880C–82880C. International Society for Optics and Photonics, 2012.

[14] F. Heide, D. Lanman, D. Reddy, J. Kautz, K. Pulli, and D. Luebke. Cascaded displays: spatiotemporal superresolution using offset pixel layers. *ACM Transactions on Graphics (TOG)*, 33(4):60, 2014.

[15] M. Hirsch, D. Lanman, G. Wetzstein, and R. Raskar. Construction and calibration of optically efficient lcd-based multi-layer light field displays. In *Journal of Physics: Conference Series*, volume 415, page 012071. IOP Publishing, 2013.

[16] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[17] Y. Hu, D. Rajan, and L.-T. Chia. Robust subspace analysis for detecting visual attention regions in images. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 716–724. ACM, 2005.

[18] F.-C. Huang, D. P. Luebke, and G. Wetzstein. The light field stereoscope. In *SIGGRAPH Emerging Technologies*, pages 24–1, 2015.

[19] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11):1254–1259, 1998.

[20] P. Jiang, H. Ling, J. Yu, and J. Peng. Salient region detection by ufo: Uniqueness, focusness and objectness. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1976–1983, 2013.

[21] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In *Computer Vision, 2009 IEEE 12th international conference on*, pages 2106–2113. IEEE, 2009.

[22] D. Lanman, G. Wetzstein, M. Hirsch, W. Heidrich, and R. Raskar. Polarization fields: dynamic light field display using multi-layer lcds. *ACM Transactions on Graphics (TOG)*, 30(6):186, 2011.

[23] D. Lanman, G. Wetzstein, M. Hirsch, and R. Raskar. Depth of field analysis for multilayer automultiscopic displays. In *Journal of Physics: Conference Series*, volume 415, page 012036. IOP Publishing, 2013.

[24] J. Li, Y. Tian, and T. Huang. Visual saliency with statistical priors. *International journal of computer vision*, 107(3):239–253, 2014.

[25] N. Li, B. Sun, and J. Yu. A weighted sparse coding framework for saliency detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5216–5223, 2015.

[26] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu. Saliency detection on light field. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2806–2813, 2014.

[27] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang. Saliency detection via dense and sparse reconstruction. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2976–2983, 2013.

[28] W. Liao, S. Wang, M. Sun, P. Surman, Y. Zheng, J. Yuan, and X. W. Sun. 19-5l: Late-news paper: Perceptually optimized dual-layer light field 3d display using a moiré-aware compressive factorization. In *SID Symposium Digest of Technical Papers*, volume 47, pages 235–238. Wiley Online Library, 2016.

[29] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum. Learning to detect a salient object. *IEEE Transactions on Pattern analysis and machine intelligence*, 33(2):353–367, 2011.

[30] A. Maimone, R. Chen, H. Fuchs, R. Raskar, and G. Wetzstein. 36.1: Wide field of view compressive light field display using a multilayer architecture and tracked viewers. In *SID Symposium Digest of Technical Papers*, volume 45, pages 509–512. Wiley Online Library, 2014.

[31] A. Maimone, G. Wetzstein, M. Hirsch, D. Lanman, R. Raskar, and H. Fuchs. Focus 3d: Compressive accommodation display. *ACM Trans. Graph.*, 32(5):153–1, 2013.

[32] Y. Pan, Y. Wang, J. Liu, X. Li, and J. Jia. Fast polygon-based method for calculating computer-generated holograms

in three-dimensional display. *Applied optics*, 52(1):A290–A299, 2013.

[33] A. Rav-Acha, Y. Pritch, and S. Peleg. Making a long video short: Dynamic video synopsis. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 435–441. IEEE, 2006.

[34] P. L. Rosin. A simple method for detecting salient regions. *Pattern Recognition*, 42(11):2363–2371, 2009.

[35] C. Shen and Q. Zhao. Learning to predict eye fixations for semantic contents using multi-layer sparse network. *Neurocomputing*, 138:61–68, 2014.

[36] X. Shen and Y. Wu. A unified approach to salient object detection via low rank matrix recovery. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 853–860. IEEE, 2012.

[37] J. Shi, L. Xu, and J. Jia. Discriminative blur detection features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2965–2972, 2014.

[38] W. Song, Q. Zhu, T. Huang, Y. Liu, and Y. Wang. 26.3: Volumetric display system using multiple mini-projectors. In *SID Symposium Digest of Technical Papers*, volume 44, pages 318–321. Wiley Online Library, 2013.

[39] Z. Tu, Z. Guo, W. Xie, M. Yan, R. C. Veltkamp, B. Li, and J. Yuan. Fusing disparate object signatures for salient object detection in video. *Pattern Recognition*, 72:285–299, 2017.

[40] R. Valenti, N. Sebe, and T. Gevers. Image saliency by isocentric curvedness and color. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2185–2192. IEEE, 2009.

[41] R. Vidal, Y. Ma, and S. Sastry. Generalized principal component analysis (gpca). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12):1945–1959, 2005.

[42] Y. Wei, F. Wen, W. Zhu, and J. Sun. Geodesic saliency using background priors. In *European conference on computer vision*, pages 29–42. Springer, 2012.

[43] G. Wetzstein, D. Lanman, W. Heidrich, and R. Raskar. Layered 3d: tomographic image synthesis for attenuation-based light field and high dynamic range displays. In *ACM Transactions on Graphics (ToG)*, volume 30, page 95. ACM, 2011.

[44] G. Wetzstein, D. Lanman, M. Hirsch, W. Heidrich, and R. Raskar. Compressive light field displays. *IEEE computer graphics and applications*, 32(5):6–11, 2012.

[45] G. Wetzstein, D. Lanman, M. Hirsch, and R. Raskar. Tensor displays: compressive light field synthesis using multilayer displays with directional backlighting. 2012.

[46] C. Yan, X. Liu, D. Liu, J. Xie, X. X. Xia, and H. Li. Omnidirectional multiview three-dimensional display based on direction-selective light-emitting diode array. *Optical Engineering*, 50(3):034003–034003, 2011.

[47] Q. Yan, L. Xu, J. Shi, and J. Jia. Hierarchical saliency detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1155–1162, 2013.

[48] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang. Saliency detection via graph-based manifold ranking. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3166–3173, 2013.

[49] J. Zhang, M. Wang, J. Gao, Y. Wang, X. Zhang, and X. Wu. Saliency detection with a deeper investigation of light field. In *IJCAI*, pages 2212–2218, 2015.

[50] Q. Zhao and C. Koch. Learning saliency-based visual attention: A review. *Signal Processing*, 93(6):1401–1407, 2013.

[51] W. Zhu, S. Liang, Y. Wei, and J. Sun. Saliency optimization from robust background detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2814–2821, 2014.