EXTENDING THE FOV FROM DISPARITY AND COLOR CONSISTENCIES IN MULTIVIEW LIGHT FIELDS

Zhao Ren, Qi Zhang, Hao Zhu, Qing Wang

School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China

ABSTRACT

Light field, which is captured by a plenoptic camera, is always limited in its narrow field of view (FOV) by the physical size of the aperture. To break through the restriction, we propose to extend the FOV using multiview light fields. A series of light fields are acquired by translating the camera at isometric spatial positions. In contrast to previous methods, our algorithm is the first that achieves light field registration and rendering based on epipolar plane image (EPI) properties, including disparity and color consistencies. Furthermore, the aliasing caused by the under-sampling in the angular space is eliminated by synthesizing novel views in the EPI space. Experimental results on the real scene data have demonstrated the effectiveness of our algorithm.

Index Terms— light field, field of view (FOV), epipolar plane image (EPI), disparity and color consistencies, novel view synthesis

1. INTRODUCTION

With the emergence of computational photography and development of high-resolution imaging sensor, light field cameras were springing up in the electronic market since 2010. In contrast to traditional cameras, light field camera captures both spatial and directional information of light rays simultaneously in a single shot, which provides a brilliant platform for most imaging applications in conventional computer vision tasks, such as depth estimation, super-resolution, saliency detection, image stitching and so on.

Rendering an image with a wide FOV, as know as a popular topic in computer vision, has been researched over the past decades. Unfortunately, image stitching results from 2D images are not satisfying due to the disparity difference [1]. Thus, light field stitching starts to address this issue on account of its capability of recording 4D light rays. To extend the FOV, current light field stitching methods are achieved by taking advantage of multiview light fields. The most state-of-the-art algorithms are based on a motion matrix between two light fields [2, 3, 4, 5]. Except that, a panorama light field can also be rendered by utilizing focal stacks [6] or minimizing the root mean square (RMS) error of the overlapped rays after
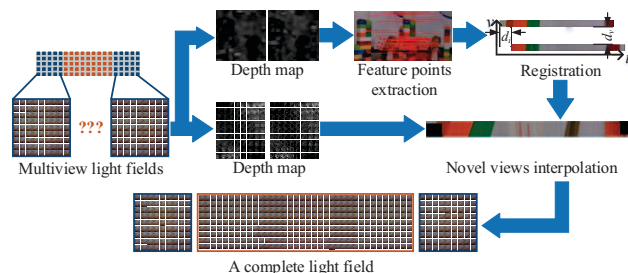
Fig. 1. The pipeline of the proposed method. Multiple novel views are interpolated between two light fields, which are captured by translating a plenoptic camera without continuous views.

the registration [7]. However, the above-mentioned methods are mostly based on a limited camera motion in order to obtain continuous sampled rays at least, which badly increases the storage burden. For instance, 36 or 24 light fields would be captured when rotating the camera at $10°$ or $15°$ per step in [7].

Hence, the algorithm on generating novel views appears to solve the problem. Novel views can be generated from two micro-baseline images [8], part of views [9] or views on four corners [10] for reconstructing a light field. Nevertheless, the existing methods are only applicable for the single-view light field which is lack of less than ten views.

In the paper, we put emphasis on interpolating novel views between two light fields under specific circumstance, as shown in Fig.1. We acquire light fields by translating light field camera with a fixed spacing, which leaves dozens of views vacant between two neighbouring light fields. Inspired by the disparity consistency on EPI [11] and color consistency of Lambertian surface [12], our approach is the first that registers light fields and interpolates dozens of views based on them.

In this paper, we make two contributions:

1) We propose the novel views interpolation to reduce the size of captured data.

2) The registration and novel views interpolation algorithms are achieved based on disparity and color consistencies on EPI, instead of utilizing overlapped rays.

ICIP 2017

2. PROBLEM DESCRIPTION

Let us introduce the basic idea of our proposed method first. We assume the light field is parameterized by a two-parallel-plane function $L(u, v, s, t)$, where (u, v) represents the angular direction and (s, t) is the spatial position.

2.1. Disparity consistency

The disparities of the points on EPI from a common 3D point would be equal theoretically because its depth keeps invariant, as shown in Fig.2(a), in which points (v_1, t_1) and (v_2, t_2) come from the same point in the 3D space so that their disparities are infinitely close. For multiview light fields which are recorded by translating the camera in a specific way (ensuring the optical center of light field camera always on a specific plane), a complete EPI can be reconstructed by captured EPIs. Thus, no matter whether the point is inside or outside the captured EPIs, its disparity is equal to the disparities of those points on its EPI tube.

2.2. Color consistency

We assume all the objects are belonging to the Lambertian surface reflection model, as shown in Fig.2(b). Thus, part of an object can be viewed as a convex hull and it is available to get the intensity I_p of the rays reflected by the convex hull surface from light source,

$$I_p = A\alpha_p \max(0, N_p \cdot V) \qquad (1)$$

where A is the strength of light source at infinity, α_p is the albedo of the p-th pixel on the object surface, N_p is the normal at the p-th surface point and V is the direction of light source. However, all pixels on the object surface share the same albedo α and the reflected intensity is independent of the normal of Lambertian surface under the Lambertian assumption. Furthermore, the light source direction V is assumed to lie in the half-sphere relative to the object surface

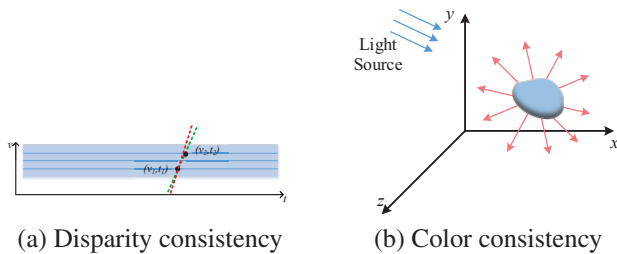(a) Disparity consistency   (b) Color consistency

Fig. 2. Illustrations of disparity and color consistencies. (a) The red and green lines represent the disparities of that two points on the EPI separately. (b) All the rays from a convex hull are assumed to share the same intensity based on the Lambertian surface assumption.

so that $N_p \cdot V > 0$. Therefore, (1) can be simplified as,

$$I_p = A\alpha N \cdot V \qquad (2)$$

where N and α are both constants.

Hence, multiview points from the same spatial point hold the same intensity no matter where the camera is located and colors (RGB values) of them are equal on sub-images or EPIs.

3. THE PROPOSED METHOD

3.1. Registration

The light fields are recorded by translating the light field camera in the proposed approach. Since the intrinsic parameter of the camera keeps unchangeable and the jitter during camera motion can be ignored, all the light fields are represented by (u, v, s, t) on a common two-plane form, as shown in Fig.3(a). Thus, computing the distance $d_i, i = 1...n - 1$ between each two light fields is the kernel task of registration.

First of all, we start by varying d_i with a fixed range $[-10mm, +10mm]$ around its initial value $D_i, i = 1...n - 1$ with a fixed step $0.2mm$. To accelerate the minimization process, it would be better to use the central view rather than the whole light field since it is evidently the clearest and sharpest one among all views. Considering feature points are more representative than other pixels, m corner point pairs are extracted by the SIFT algorithm [13] from each sub-image pair on the central view. Next, the cost function of feature points is minimized according to the disparity consistency,

$$E_{reg}(d_i) = \sum_{j=1}^{m} (|disp_l(j) - k(d_i, j)| + |disp_r(j) - k(d_i, j)|). \quad (3)$$

It is noticed that disparities $disp_l(j)$ and $disp_r(j)$ of the j-th point in the left and right sub-images are computed by the method in [8] respectively. Additionally, $k(d_i, j)$ is the reciprocal of the slope of red line in Fig.3(b).

3.2. Novel views interpolation

As our captured light fields are lack of several views, it is essential to interpolate the missing views, as shown in Fig.5. Otherwise, aliasing will be brought by the insufficient angular sampling [14].

Accurate disparities for all the objective pixels are needed to improve the accuracy of interpolation. To compute an initial disparity value for each interpolated pixel before minimization, we shall draw support from edges in the central view to divide pixels into different regions, which can be detected by 'Canny' operator. It is feasible to use disparities of these points to draw some initial tubes on the EPI. Pixels which fall into the area constructed by the tubes can be initialized in an interval disparity (disparities of the left and right tube in the region). Thus, the minimization progress must be
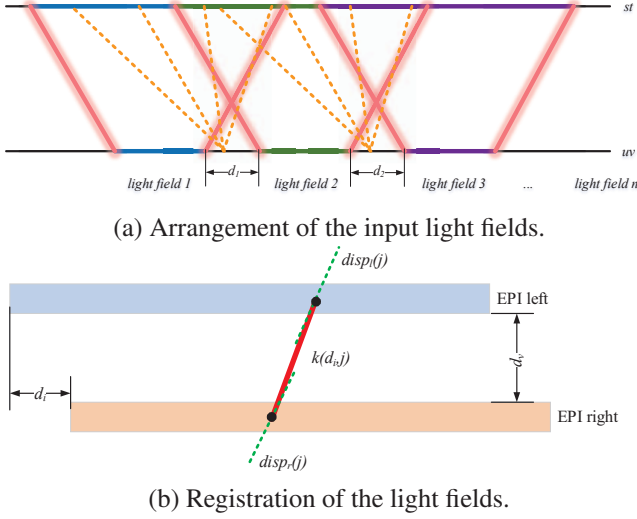
(a) Arrangement of the input light fields.

(b) Registration of the light fields.

Fig. 3. (a) Totally n light fields are represented on the common two-plane $uv-st$. The pink lines are light rays captured by the camera and the orange ones are missing rays. (b) The blue EPI is from the left light field and the orange one is from right. Vertical distance between two EPIs d_v is relative to d_i. $disp_l(j)$ and $disp_r(j)$ (green lines) are coherent with geometry slope $k(d_i, j)$ (red line).

faster because the disparity of each pixel is iterated in its own interval.

The disparity is varied in a range of $[-DP, +DP]$ with a step of $d_{step} = 0.01$. First, the EPI is refocused [15] to facilitate computing the objective function when the point is focused on during the minimization. For a point (s^*, t^*), it is refocused on Rf such that

$$Rf(s^*, t^*) = Epi(s^*, t) = Epi(s^*, disp*(s-s^*)+t^*), \quad (4)$$

where $disp = -DP + m*d_{step}, m = 0, ..., \frac{2DP}{d_{step}}$.

Based on the disparity consistency and color consistency, the cost function for a pixel (s^*, t^*) is minimized by,

$$E_{rend}(disp(s^*, t^*)) = E_c + E_d, \quad (5)$$

where

$$E_c = \sqrt{\sum_{c=1}^{3} \sum_{w=-5}^{5} (\sum_{i=1}^{view_v} (Rf_l(l_i, t^*+w) - Rf_r(r_i, t^*+w)))^2} \quad (6)$$

and

$$E_d = \alpha|\overline{disp}_l(t^*) - \overline{disp}_r(t^*)| + \\ |\overline{disp}_l(t^*) - disp(s^*, t^*)| + |\overline{disp}_r(t^*) - disp(s^*, t^*)| \quad (7)$$

where $view_v$ is the number of views along the v axis, $\overline{disp}_l(t^*)$ is the average of disparities of the pixels (s, t^*), $s = 1, ..., view_v$ and $\overline{disp}_r(t^*)$ is the average of the right one.

Besides α is ranged in $\alpha \in (0, 2)$. Thus, the color of pixel (s^*, t^*) can be interpolated by,

$$Epi(s^*, t^*) = \frac{1}{2}\left[\sum_{i=1}^{view_v} W(i)Rf(l_i, t^*) + \sum_{i=1}^{view_v} W(i)Rf(r_i, t^*)\right] \quad (8)$$

where **W** is a weight vector to balance the contribution of different views, among which the pixel on the middle view has the biggest weight as it is the sharpest one, defined as

$$W(i) = \begin{cases} \frac{i}{\sum_{i=1}^{view_v} W(i)}, & i \in [1, \lceil \frac{view_v}{2} \rceil] \\ \frac{view_v - i + 1}{\sum_{i=1}^{view_v} W(i)}, & i \in [\lceil \frac{view_v}{2} \rceil, view_v] \end{cases} \quad (9)$$

As the occlusion exists during camera motion, it is necessary to handle it while interpolating views. But the state-of-the-art depth estimation method can not be utilized to deal with the issue as the translation distance is too long. Therefore, we propose a judgement to interpolate the occluded pixels. For the pixel (s^*, t^*), the first judgement is designed based on the color property by the difference of average of both refocused EPIs with a predefined threshold T,

$$f(s^*, t^*) = \begin{cases} 1, & \delta_{color} \geq T \\ 0, & \delta_{color} < T \end{cases} \quad (10)$$

where $\delta_{color} = |\overline{Rf}_l(t^*) - \overline{Rf}_r(t^*)|$.

Thus, it could be obtained that the pixel is occluded when $f(s^*, t^*) = 1$. Next, its color is interpolated as follows based on the disparity property when it is occluded,

$$Epi(s^*, t^*) = \begin{cases} \sum_{i=1}^{view_v} W(i)Rf(r_i, t^*), & \delta_{disp} \geq 0 \\ \sum_{i=1}^{view_v} W(i)Rf(l_i, t^*), & \delta_{disp} < 0 \end{cases} \quad (11)$$

where $\delta_{disp} = |disp - \overline{disp}_l(t^*)| - |disp - \overline{disp}_r(t^*)|$ and $W(i)$ is defined in (9).

4. EXPERIMENTS

We employ a Lytro camera to evaluate the proposed approach. With $1.5\times$ zoom, the intrinsic parameters are calibrated, among which the size of uv plane is $4.4mm \times 4.4mm$ with 9×9 views, the size of st plane is $264.3mm \times 264.3mm$ with 328×328 pixels, and the distance between uv and st plane is $575.9mm$.

Our interpolation approach are validated as in Fig.4. We start from the same camera position and capture 3 light fields, in which the first and the last one are used as the input light fields. Our interpolated results are compared with the central view of the $2nd$ light field by the PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index Measure) values.
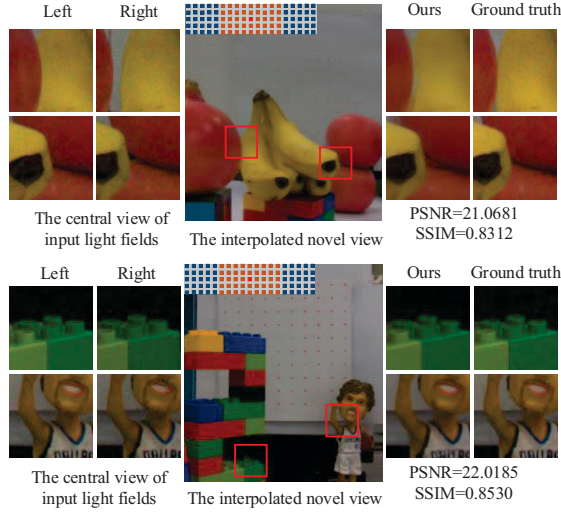
1159

Fig. 4. Comparisons of the interpolated view between two light fields at an initial distance of $20mm$ with the ground truth view between the two input light fields.

Next, we add an additional light field to the two light fields in Fig.5 and refocus the three light fields to exhibit a complete light field with a wider FOV, as shown in Fig.6. As we can see from the results, light fields are registered together and the sampling rate is enough to avoid aliasing. The battledore is not continuous when it is defocused, but it is connected well when focused on. The aliasing is caused by the low sampling rate when our interpolation result is not an absolutely complete light field, where some points are sampled by both input light fields and interpolated novel views, but some are only sampled by the input light fields.

To evaluate the effectiveness and the sampling rate difference further, a complete light field is generated from five light fields, as shown in Fig.7. The aliasing in the refocused images disappears when the objects are focused on.

5. CONCLUSION

In this paper, we present a novel approach to extend the FOV of light field based on the translation which allows capturing light field without continuous views. This is the first time to achieve registration and novel views interpolation based on EPI properties. However, there are also some limitations in our method. First, to validate the effectiveness of our approach intuitively, we simplify the camera motion to translation. Second, the proposed approach is only adaptive to Lambertian surface.

In the future, we plan to extend our approach to capture light field by rotation and translation at the same time. Furthermore, we will attempt to improve the universality of our approach by exploiting properties of Non Lambertian surface (mirror, transparent object, etc.).
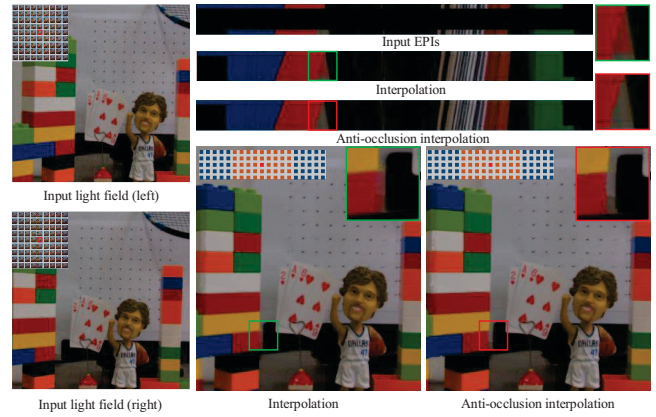

Fig. 5. Interpolation results of EPI and novel views are both presented before and after occlusion processing. The sub-images are extracted from the central interpolated views.
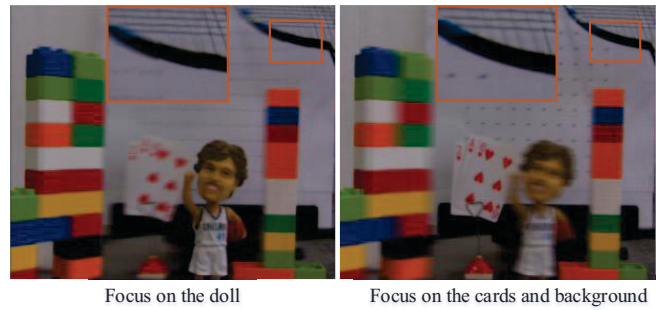

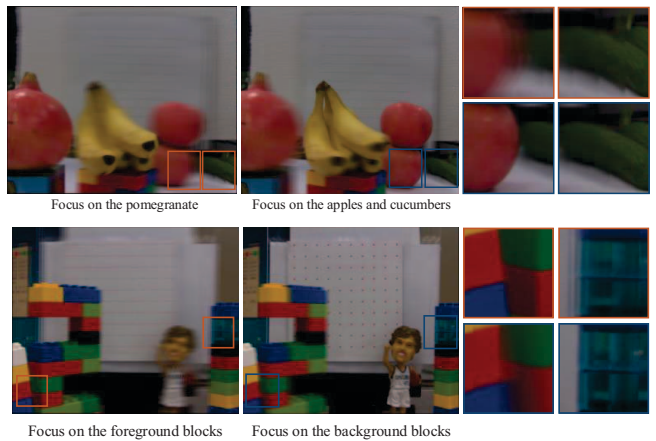Fig. 6. The refocused images on different depth planes using three light fields.

Fig. 7. The images are refocused from five light fields at initial $10mm$ intervals.

1160

6. REFERENCES

[1] Fan Zhang and Feng Liu, "Parallax-tolerant image stitching," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3262–3269.

[2] Zhou Xue, Loic Baboulaz, Paolo Prandoni, and Martin Vetterli, "Light field panorama by a plenoptic camera," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2014, pp. 90200S–90200S.

[3] Xinqing Guo, Zhan Yu, Sing Bing Kang, Haiting Lin, and Jingyi Yu, "Enhancing light fields through ray-space stitching," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 7, pp. 1852–1861, 2016.

[4] Ole Johannsen, Antonin Sulc, and Bastian Goldluecke, "On linear structure from motion for light field cameras," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 720–728.

[5] M Umair Mukati and Bahadir K Gunturk, "Light field registration: A multi-view geometry approach," in *Signal Processing and Communication Application Conference (SIU), 2016 24th*. IEEE, 2016, pp. 1289–1292.

[6] Clemens Birklbauer, Simon Opelt, and Oliver Bimber, "Rendering gigaray light fields," in *Computer Graphics Forum*. Wiley Online Library, 2013, vol. 32, pp. 469–478.

[7] Clemens Birklbauer and Oliver Bimber, "Panorama light-field imaging," in *Computer Graphics Forum*. Wiley Online Library, 2014, vol. 33, pp. 43–52.

[8] Zhoutong Zhang, Yebin Liu, and Qionghai Dai, "Light field from micro-baseline image pair," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3800–3809.

[9] Lixin Shi, Haitham Hassanieh, Abe Davis, Dina Katabi, and Fredo Durand, "Light field reconstruction using sparsity in the continuous fourier domain," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 1, pp. 12, 2014.

[10] Nima Khademi Kalantari, Ting-Chun Wang, and Ravi Ramamoorthi, "Learning-based view synthesis for light field cameras," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 6, pp. 193, 2016.

[11] Sven Wanner and Bastian Goldluecke, "Globally consistent depth labeling of 4d light fields," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 41–48.

[12] Mohammadul Haque, Avishek Chatterjee, Venu Madhav Govindu, et al., "High quality photometric reconstruction using a depth camera," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2275–2282.

[13] D. G. Lowe, "Object recognition from local scale-invariant features," in *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1999, p. 1150.

[14] Zhaolin Xiao, Heng Yang, Qing Wang, and Guoqing Zhou, "Depth free entropy based ghost removal on synthetic aperture images," in *Signal Processing, Communication and Computing (ICSPCC), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1–5.

[15] Ren Ng, *Digital light field photography*, Ph.D. thesis, stanford university, 2006.