None Ghosting Artifacts Stitching Based on Depth Map for Light Field Image

Wenyuan Zhang, Shengyang Zhao, Wei Zhou, and Zhibo Chen(✉)

CAS Key Laboratory of Technology in Geo-spatial Information Processing and Application System
University of Science and Technology of China, Hefei 230027, China
chenzhibo@ustc.edu.cn

Abstract. Due to hardware limitations, existing light field (LF) capturing devices cannot offer sufficient field of view for building 6 degrees of freedom (6 DOF) VR applications. LF image stitching methods can be used to address this problem. The state-of-the-art LF stitching methods highly depend on the stitching accuracy of center view which is essentially a 2D image stitching task. However, conventional 2D image stitching methods usually suffer from the ghosting artifacts. In this paper, a None Ghosting Artifacts (NGA) stitching method is proposed to tackle this problem. We theoretically reveal the intrinsic cause of the ghosting artifacts and then further verify that different depth scenes require different homography matrices for warping. Therefore, the clustered depth map is employed to segment the scene into several layers, and the layer-specific homography matrix is computed for warping. An interpolation mechanism is also proposed to ensure that each layer has its own transformation. Compared with state-of-the-art stitching methods aiming to alleviate ghosting artifacts, experimental results show that the proposed method not only stitches images without ghosting artifacts, but also achieves realistic perspective transformation.

Keywords: Ghosting artifacts · Depth map · Image stitching · Light field stitching.

1 Introduction

A light filed (LF) consists of a large collection of rays that store radiance information in both spatial and angular dimensions [6]. With extra angular light information, the LF can provide 6 degrees of freedom (6 DOF) experience. It is considered as a promising technique for future immersive multimedia applications, such as 3D TV and virtual reality (VR) applications.

LF can be captured by camera arrays or lenslet LF cameras, e.g. Lytro [9]. Both the camera array and lenslet camera cannot provide sufficient filed-of-view (FOV) for building the 6 DOF VR applications. In order to introduce the LF techniques into the market successfully , it is urgent to stitch the LF captured by multiple perspectives. A 6 DOF panoramic VR scene can be constructed by stitching the sub-views of multiple LF images.

Many research works have been carried out for LF stitching. [6, 7, 1] represent the LF as a 4D function and extend 2D image stitching algorithms to stitch LF in 4D dimension. However, they suffer from the high complexity caused by high data dimensions. [2] essentially stitches the LF in the frequency domain, which converts 4D plenoptic data to 3D focal stack before stitching and thus is limited to Lambertian scenes. The state-of-the-art method [3] represents the LF images as views array and stitch the center view with 2D stitching method. Then the stitching operation is propagated to other views. The method is based on propagation and robust for different scenes and devices.However, they suffer from the ghosting artifacts, which is an intrinsic drawback of conventional 2D image stitching. Solving the ghosting artifact is our motivation for this paper.

When the captured scene contains a relatively larger depth range, there will exist a large parallax between the adjacent images. In this case, traditional 2D image stitching will result in severe ghosting artifacts. Despite blending and feathering methods are exploited for to de-ghost, a good initial stitching can not only extremely improve the results, but also impose a much lower requirement on subsequent de-ghosting and post-processing, especially for the large parallax condition. [10, 5, 8, 11] all aimed to optimize the initial stitching, but none of them attempted to investigate the essence of the ghosting artifacts and thoroughly solved this problem.

In this paper, we first prove that the essential cause of the ghosting artifacts is that pixels at different depth planes need their own transformation matrix. Fortunately the depth information can be estimated from the LF image. Therefore, we propose a none ghosting artifacts (NGA) stitching method based on the depth map. We first use the well-known k-means algorithm to cluster the depth map of central view. Then we segment the scene into several layers with the depth map. Finally, different scene layers are transformed with their own homographies, which are computed by the matching feature points belonging to each depth plane with RANSAC algorithm. Considering that some depth planes may not have enough matching points to obtain an accurate enough transformation, we use a camera parameter based interpolation to ensure each layer has its own homography matric.

Our contributions is in the following three aspects:

1). We theoretically point out the reason of ghosting artifact;

2). We propose the NGA stitching algorithm for the center view of LF images;

3). We propose an interpolation mechanism to make NGA more robust.

The rest of the paper is organized as follows: Section 2 surveys related work. Sec-tion 3 introduces our proposed method NGA. Experimental results are presented in Section 4, and we conclude our work in Section 5.

2 Related work

Our work is related to LF stitching and 2D image stitching. In this section, we review both them in the follows.

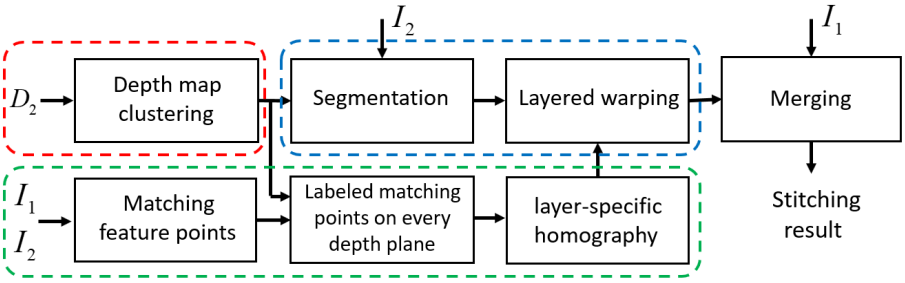

Fig. 1: Framework of the NGA stitching model. I_1 and I_2 denote the central view of two LFs respectively. D_2 denotes the depth map of I_2. The red box in the figure represents depth map clustering, the green box represents homography calculation, and the blue box represents layer-specific transformation.

2.1 LF stitching

As mentioned in section 1, we divide the existing LF stitching algorithms into three main categories as follows:

Stitching in ray space. This category [6, 7, 1] considers the transformation of LF images in the ray space. [6, 7] first find the matching rays, and then use a 5x5 transformation matrix to stitch LF images. These methods have a large amount of calculation due to the high dimensionality of the ray transformation matrix. [1] proposes to adopt the re-parameterization with double cylinders to handle different complex scenes. However, it requires highly accurate camera calibration and camera control, which is hard to satisfy in practice.

Stitching with focal stack. This category [2] generates panorama LF by first converting input LF images into focal stacks. Then they stitch these focal stacks and convert the resulting panorama focal stack into a LF using linear view synthesis. These method will fail in non-Lambertian scenes which loses the advantages of light-field imaging in general.

Stitching based on multi-view. The methods based on multi-view represent the LF images as view arrays. A general framework is established in [3], which propagates arbitrary spatial deformations operated in the center view to all other perspective views consistently. Once this method is used to stitch LF image, which may first stitch the center view with 2D stitching method, and then propagate the operation to other views without destroying the disparity consistency.

By comparing existing LF stitching algorithms, it shows that[3] has faster stitching speed and has a lower requirement for camera control accuracy. Obviously, such method will be highly affected by the 2D stitching. Thus, It is desirable to optimize the 2D image stitching.

2.2 2D image stitching

A classic 2D stitching model was proposed in [4], which first extracts feature points, then matches the feature points, and finally calculates a global homography to fit the matched feature points. However, this model have two limited conditions: (1) the scene of image is planar or (2) the views differ purely by rotation [10]. But in practice, the two assumptions are hard to satisfy, and hence there will generate obvious ghosting artifacts and misalignment due to parallax. By assuming that the scene contains a ground plane and a distant plane, [5] proposed dual homography warps for image stitching. Essentially [5] is a special case of a piece-wise projective warping, which is more flexible than using a single homography. But it is not robust for complex scenes. In [8], multiple affine transformations are used to make more precise local alignment. [10] replaces local affine transformations in [8] with local homography transformations, which makes local adaptation much better. A simple moving Direct Linear Transformation (DLT) method in [10] is used to estimate the local parameters, by providing higher weights to closer feature points and lower weights to the farther ones. However, once the local mesh is at the boundary of the objects, it is impossible to ensure those closer feature points belong to the same depth plane. For better stitching, Content-preserving warps (CPW) in [11] combine the global transformation with the block based warping. Therefore,those methods can only alleviate but not absolutely avoid the ghosting artifacts. In the next section, we point out the essential reason for ghosting artifacts and shows that the proposed NGA can solve this problem.

3 None Ghosting Artifacts Stitching

In this section, we first point out that each depth plane corresponds to a specific homography matrix. That is why the conventional algorithms always produce ghosting artifacts. Further, we derive the relationship between the homographies of different depth planes. Based on these analysis, we proposed our NGA stitching method.

3.1 3D Warping Function

The relationship between a 3D point $\mathbf{M}$ and its image projection $\mathbf{m}$ is given by [12].

$$s\tilde{\mathbf{m}} = \mathbf{ARM} + \mathbf{At} \ , with \ \mathbf{A} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{1}$$

where s is the depth of a 3D point, and $\mathbf{A}$ is the intrinsic matrix of the camera. $\mathbf{R}$ and $\mathbf{t}$ denote the camera rotation and translation, respectively. We use $\sim$ to denote the homogeneous coordinates of $\mathbf{m}$.

We can consider two cameras conditions. then the projection of point $\mathbf{M}$ are denoted by $\mathbf{m}_1$ and $\mathbf{m}_2$. The relationship between them can be derived as [12]

$$\mathrm{s}_2\tilde{\mathbf{m}}_\mathbf{2}=\mathrm{s}_1\mathbf{A}\mathbf{R}\mathbf{A}^{-\mathbf{1}}\tilde{\mathbf{m}}_\mathbf{1}+\mathbf{A}\mathbf{t}. \tag{2}$$

When capturing the LF panorama, the camera may have two kinds of movement, i.e. the translation and rotation. To simplify the derivation, we separate the two conditions. If we consider translating only, there should be $s_2 = s_1 = s$, $\mathbf{R} = \mathbf{E}$,where $\mathbf{E}$ is an indetity matrix. the equation (2) can be simplified as

$$\tilde{\mathbf{m}}_\mathbf{2} = \tilde{\mathbf{m}}_\mathbf{1} + \frac{\mathbf{A}\mathbf{t}}{s}, \tag{3}$$

where $\mathbf{t} = (t_x, t_y, 0)^T$. The equation (3) can also be written as

$$\tilde{\mathbf{m}}_\mathbf{2} = \mathbf{H}\tilde{\mathbf{m}}_\mathbf{1} \ , with \ \mathbf{H} = \mathbf{E}+\frac{\mathbf{A}\mathbf{t}}{s}, \tag{4}$$

It shows that $\mathbf{H}$ in equation (4) is only decided by the depth. Thus, it is reasonable to use different homography for different depth layers. Further, the relationship between the homographies of different depth planes can be expressed as

$$\mathbf{H}_\mathbf{2} = \frac{s_1}{s_2}(\mathbf{H}_\mathbf{1} - \mathbf{E}) + \mathbf{E}, \tag{5}$$

where s_1 and s_2 are different depth values, and $\mathbf{H}_1$ and $\mathbf{H}_2$ is the homgraphy correspond to s_1 and s_2, respectively.

If only the rotation is considered, we have $\mathbf{t} = (0, 0, 0)^T$, $s_2 \approx s_1$, the equation (2) can be simplified as

$$\tilde{\mathbf{m}}_\mathbf{2} = \mathbf{H}\tilde{\mathbf{m}}_\mathbf{1} \ , with \ \mathbf{H} = \mathbf{A}\mathbf{R}\mathbf{A}^{-\mathbf{1}}. \tag{6}$$

We can find that $\mathbf{H}$ is not related to depth in this case. However, actually a pure rotation is nearly impossible because the main lens cannot be treated a point. This means the second term in equation (3) will affect the translation.

Now we conclude that if we want to align two projection pixel, we need to calculate the transformation based on the depth value, and different depth plane should have its own matrix. This is the reason that conventional work always produce misalignment and introduce ghosting artifacts.

3.2 Component of NGA

Based on the above analysis, we propose the None Ghosting Artifacts (NGA) stitching model. As shown in Fig. 1, the proposed model consists of three key parts : depth plane clustering, homography calculation, and layer-specific transformation.

Depth plane clustering. According to the theoretical derivation, each different value in the depth map corresponds to a different homography. However, calculating homography for all depth values is time-consuming. Also for some

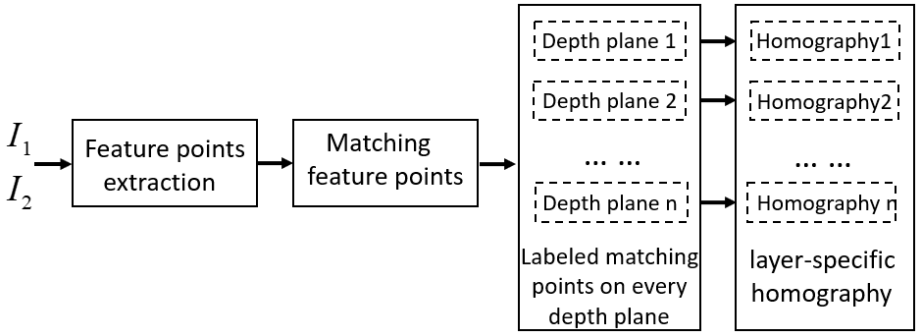

Fig. 2: The homography calculation.

depth values there may be not enough feature points for calculation. Moreover the estimated depth maps are usually noisy. Therefore, we cluster the depth map by K-means algorithm. The K value is determined by Calinski-Harabasz criterion.

Homography calculation. This part is shown in Fig. 2. For input images I_1 and I_2, the SIFT feature points are extracted and matched. For each depth layer, we count the feature points locating within it. If the num of feature points is more than a threshold N_fp , we calculate its own homography transformation by RANSAC algorithm; otherwise we obtain the transformation matric by our interpolation method, which is detailedly described in Section 3.3.

Layer-specific transformation. Due to I_2 is transformed to I_1,we first segment I_2 into several layers by the clustered depth map. Then we perform the corresponding transformation on each layer. The layers are merged by depth order, i.e. the farthest layer is put first and so on. It is noticeable that the post-processing methods, like blending or feathering can be used for any other algorithms. Thus, we do not adopt them to help fairly comparing.

The process of Layer-specific transformation is shown in Fig. 3. Due to the segmentation and layering operations, each depth plane monopolizes a different transformation. Thus the stitching results can achieve realistic perspective conversion.

3.3 Interpolation module

The significance of this module is to make up for the unstability of the homography calculation in the NGA framework under certain circumstances. There are two special cases that make it hard to obtain accurate homographies of some depth layers. On the one hand, the matched feature points exist only in the overlapping areas of the I_1 and I_2, which can result in the fact that some depth planes may not have matched feature points, such as the non-overlapping area in I_2 exists a depth plane that does not exist in the overlapping area, so that homography of this depth plane cannot be calculated. On the other hand, some

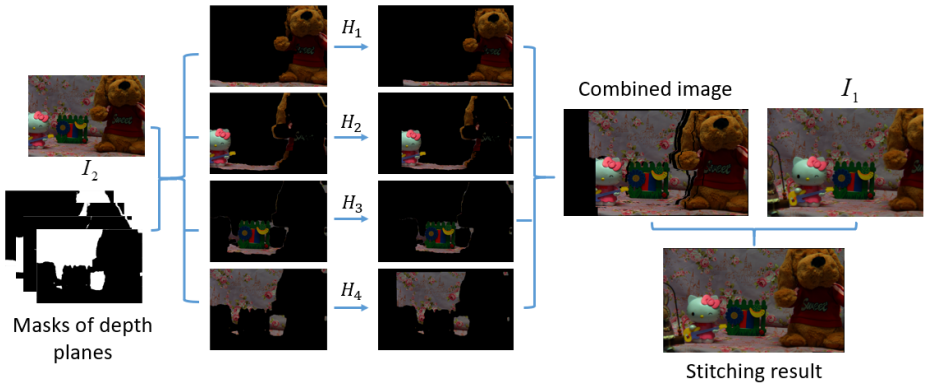

Fig. 3: The layered transformation diagram.

depth planes contain fewer matching feature points so that robust homography cannot be obtained. To solve this problem, we have adopted an interpolation mechanism. Formula (5) in Section 3.1 shows that we can only use a known homography to compute an unknown homography. In order to get a more robust result, we use all known homgraphy to compute the unknown homography of the same depth plane. Thus we can get multiple candidate homographies that are calculated as

$$\mathbf{H}_*^i = \frac{s_k^i}{s_*}(\mathbf{H}_k^i - \mathbf{E}) + \mathbf{E}, \tag{7}$$

where subscript k present the meaning of known, so s_k^i and s_* are the depth value of the i-th known homography and the depth value of the unknown homography respectively. $\mathbf{H}_k^i$ is the i-th known homography, and $\mathbf{H}_*^i$ is the homography computed by the i-th known homography. The unknown homography is estimated from the weighted homography as

$$\mathbf{H}_* = \mathbf{E} + \sum_{i=1}^{n} w_*^i(\mathbf{H}_*^i - \mathbf{E}). \tag{8}$$

The weights $\{w_*^i\}_{i=1}^n$ change according to the distance of between the i-th depth layer of known homography and the depth layer of unknown homography, which are calculated as

$$w_*^i = \exp(-\|s_k^i - s_*\|^2/\sigma^2). \tag{9}$$

Here, σ is a scale parameter. Therefore, we can use known and stable homographies to calculate those unstable homography by equation (8).

4 Experiment Results

In this section, we carry out experiments to verify the proposed stitching methods. First we compare our algorithm with state-of-the-art methods [10, 11, 4].

We analyze the ghost artifact and check whether the stitching methods offer a true perspective transformation. The results demonstrate that our method outperforms all the others in both two aspects. To validate the proposed interpolation method, we also offer the stitching results and compare it with the APAP method.

Fig. 4: The qualitative comparisons of alleviating ghosting artifacts. List of acronyms and initialisms: global homography, as-projective-as-possible, parallax-tolerant image stitching.

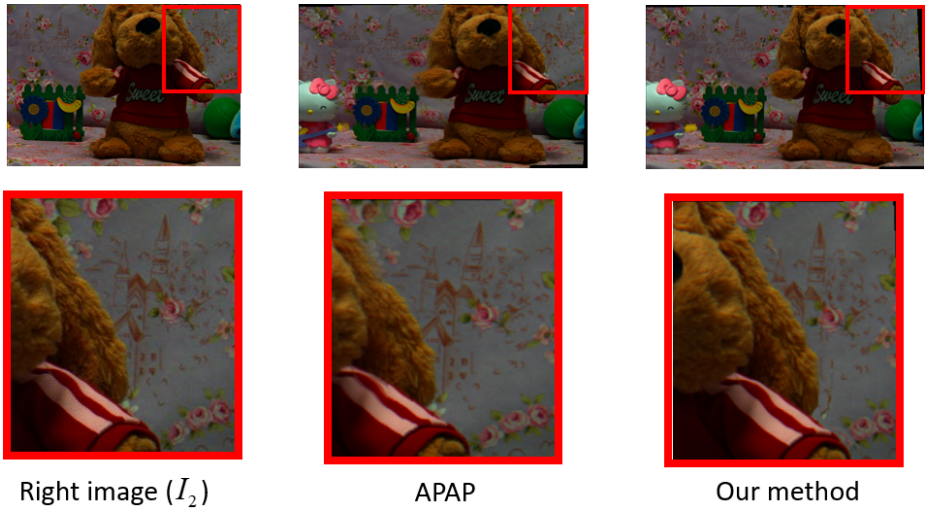

Fig. 5: Comparison of perspective conversion. comparing APAP stitching result with I_2, the relative position of the background layer and the bear did not change. Comparing our stitching result with I_2, the background moved relative to the bear to the left.

The test LF images are captured by Lytro illum. The Lytro camera is fixed in a slide.Only translational motion, without any rotational movement, is included in the experiments. The raw data is then decomposed by the LF Toolbox 0.4 into view arrays of the dimensions 15x15x434x625. The center view depth map is estimated by the Lytro desktop software.

4.1 Comparisons with other methods

An adequate stitching algorithm should transform the images captured from different perspectives into one coordinate system and merge them together. Thus the algorithm should offer a true perspective transformation and introduce as few ghosting artifacts as possible.

We here compare our method against other state-of-the-art stitching methods : as-projective-as-possible (APAP) [10], parallax-tolerant image stitching (PA-TOIS) [11] and the global homography (GH) [4] as baseline. To fairly compare the methods, we do not consider sophisticated post-processing like seam cutting and straightening, and simply blend the aligned images by intensity averaging so that any misalignments remain conspicuous.

As shown in Fig. 4, the GH method produces severe ghosting artifacts. It is easy to be understood because a global homography cannot satisfy all the pixels lying on different depth planes, as what we emphasize in Section 3. The APAP divide the image into blocks and calculate transformation for each block. However if the block lies cross the edge of the objects, APAP will fail because

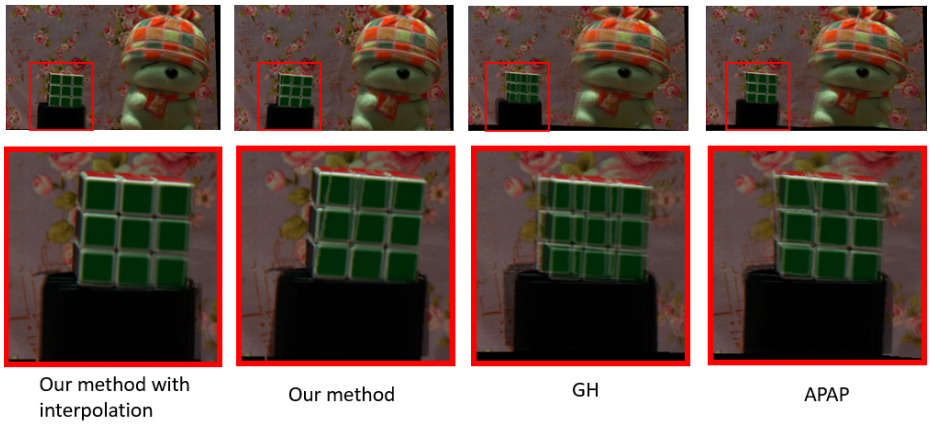

Fig. 6: Interpolation result. The homography of second depth plane (Rubik's Cube) is interpolated by weighted homographies of first depth plane (MashiMaro) and third depth plane (background).

it perform the same transformation on pixels belong to different depth planes. Thats why it works well at the center of the object (like the pen container in Fig. 4) but fails at the object boundary (e.g. the teddys hand and hello kitty head in Fig. 4). The PATOIS employs a block-based warping. Therefore, it suffers from the object boundary for the reason we give above. Our method obviously outperforms all the others. Almost no ghost artifacts or blur could be found. It verifies that our analysis for the essential cause of ghosting artifacts is reasonable and our proposed method really works.

Secondly we check if these methods provide the true perspective transformation. After transforming the right image(I_2) to the left one(I_1), the aligned image should show the perspective from the left camera. Objects at different depth have different disparity. Thus their relative position should be changed by a true perspective transformation. But we may see, for the methods shown in Fig. 5, only our method can achieve this.

4.2 Interpolation module validation

To validate our interpolation module, we use our interpolation mechanism to interpolate the homography matrixs. As shown in Fig. 6, the stitching result with our method suffer from obvious ghosting artifacts at the depth plane of Rubik's Cube. The reason for this phenomenon is that the depth plane of Rubik's Cube contains fewer matching feature points, which results in the inaccuracy of Homography calculated using those points. At the same time, compared with GH and APAP, the stitching result with interpolation produces a surprising result. This shows that our interpolation mechanism can achieve better results when the homographies of some depth planes cannot be calculated or calculated accurately by using the feature points directly.

Table 1: Comparison of the running time of each method.

	GH	APAP	NGA
feature points extraction	0.69s	0.69s	0.69s
RANSAC algorithm	9.89s	9.89s	9.89s
compute homographies	0.0101s	1.0924s	0.0136s
total	23.0s	24.4s	23.2s

4.3 Run time

We implemented all methods in MATLAB2016 and on a DESKTOP-40REPQJ computer running at 3.40GHz. The obvious difference between these methods(GH, APAP and NAG) is the number of times the homographies are computed, but it doesn't take much time to finish this step. In fact, it takes most of the time to extract feature points and use RANSAC algorithm in 2D image stitching, so the total running time of these methods is about the same. As shown in Table 1, the experimental result is consistent with our conjecture.

5 Conclusion and Future Work

We have proposed a simple but effective stitching method NGA to stitch images without ghosting artifacts. We theoretically revealed the intrinsic cause of the ghosting problem and demonstrated that different depth scenes require different homographies for warping. An interpolation mechanism is also proposed to ensure that each depth plane has its own homography. the experiments results show that our proposed method not only stitch images without ghosting artifacts, but also achieve realistic perspective transformation compared with previous methods, and the interpolation mechanism we proposed is also reasonable.

We plan to implement LF stitching based on our method. We can use the parallax relationship between the off-center perspectives and the central view in LF to find areas where the off-center perspectives are the same depth planes as the central view, and to propagate homographies at various depth levels of the central perspective to the off-center perspectives.

Acknowledgements. This work was supported in part by the National Key Research and Development Program of China under Grant No. 2016YFC0801001, the National Program on Key Basic Research Projects (973 Program) under Grant 2015CB351803, NSFC under Grant 61571413, 61632001,61390514, and Intel ICRI MNC.

References

1. Birklbauer, C., Bimber, O.: Panorama light-field imaging. In: Computer Graphics Forum. vol. 33, pp. 43–52. Wiley Online Library (2014)

2. Birklbauer, C., Opelt, S., Bimber, O.: Rendering gigaray light fields. In: Computer Graphics Forum. vol. 32, pp. 469–478. Wiley Online Library (2013)
3. Birklbauer, C., Schedl, D.C., Bimber, O.: Nonuniform spatial deformation of light fields by locally linear transformations. ACM Transactions on Graphics (TOG) **35**(5), 156 (2016)
4. Brown, M., Lowe, D.G.: Automatic panoramic image stitching using invariant features. International journal of computer vision **74**(1), 59–73 (2007)
5. Gao, J., Kim, S.J., Brown, M.S.: Constructing image panoramas using dual-homography warping. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. pp. 49–56. IEEE (2011)
6. Guo, X., Yu, Z., Kang, S.B., Lin, H., Yu, J.: Enhancing light fields through ray-space stitching. IEEE transactions on visualization and computer graphics **22**(7), 1852–1861 (2016)
7. Johannsen, O., Sulc, A., Goldluecke, B.: On linear structure from motion for light field cameras. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 720–728 (2015)
8. Lin, W.Y., Liu, S., Matsushita, Y., Ng, T.T., Cheong, L.F.: Smoothly varying affine stitching. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. pp. 345–352. IEEE (2011)
9. Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., Hanrahan, P.: Light field photography with a hand-held plenoptic camera. Computer Science Technical Report CSTR **2**(11), 1–11 (2005)
10. Zaragoza, J., Chin, T.J., Brown, M.S., Suter, D.: As-projective-as-possible image stitching with moving dlt. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. pp. 2339–2346. IEEE (2013)
11. Zhang, F., Liu, F.: Parallax-tolerant image stitching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3262–3269 (2014)
12. Zhang, Z.: A flexible new technique for camera calibration. IEEE Transactions on pattern analysis and machine intelligence **22**(11), 1330–1334 (2000)