Communications in Information and Systems
Volume 17, Number 3, 171–191, 2017

A novel rendering approach for unstructured light field interpolation

Mandan Zhao and Xiangyang Hao

In this paper, a novel framework is proposed for light field reconstruction from a sparse set of views. The light field is the unstructured, and has the large disparity. We indicate that the reconstruction can be efficiently modeled as angular restoration on epipolar plane image (EPI). Based on "blur-restoration-deblur" framework, we analyze the EPI with Fourier transform. What's more, a novel rendering approach is presented by combining proposed "blur-restoration-deblur" framework and depth information to handle large disparity and unstructured light field. We evaluate our approach on several datasets and demonstrate the high performance and robustness of the proposed framework compared with the state-of-the-arts algorithms.

1. Introduction

Light field imaging [17, 23] is one of the most extensively used method for capturing the 3D appearance of a scene. Rather than a limited collection of 2D image, the light field camera is able to collect not only the accumulated intensity at each pixel but also light rays from different directions. Early light field cameras, such as multi-camera arrays and light field gantries [37], required expensive custom-made hardware or time-consuming capturing process.

In recent years, the introduction of commercial and industrial light field cameras such as Lytro [1] and RayTrix [2] have taken light field imaging into a new era. These plenoptic (light field) cameras are composed by microlens array and has the capacity of simultaneous capture. Unfortunately, due to restricted sensor resolution, they must make a trade-off between spatial and angular resolution, i.e., one can obtain dense sampling images in the spatial dimensions but sparse sampling in the angular (viewing angle) dimensions, or vice versa.

To solve this problem, some learning-based methods [14, 20, 39] are proposed to super-resolve the light field in angular dimensions using a small

set of views with high spatial resolution. Rather than some conventional researches [27, 32, 42] that focus on novel view synthesis or reconstruction of plenoptic function, the learning-based methods train the network by directly minimizing the error between the synthetic view and the ground truth image. However, the network training is data-dependent and can not be easily transferred to data with different appearance properties, which limits the universal usage of the network.

Among these methods, Wu *et al.* propose a novel learning-based framework to reconstruct high angular resolution light field on epipolar plane image (EPI). They achieve the state-of-the-art results. By taking advantage of the special structure of the EPI, the light field reconstruction can be effectively modeled as learning-based angular detail restoration on this 2D structure. Compared with the sub-aperture images, the light field data share similar property in the EPI domain.

Based on Wu *et al.*'s pipleline, we analyze the information asymmetry of the EPI and demonstrate the resulting ghosting effects in the Fourier domain. And the efficacy of the proposed "blur-restoration-deblur" framework is also validated in the Fourier domain; Second, beside the application for depth enhancement which was shown in the preliminary version, we extend the proposed framework for more applications including interpolation for unstructured input such as unstructured light fields and video sequences; Third, we present a novel rendering scheme that seamlessly combines the proposed "blur-restoration-deblur" framework and depth information to address the interpolation with large disparity. This application inherits the rendering capability of handling large disparity from depth image-based rendering technique as well as the robustness to depth uncertainties, occlusion regions and non-Lambertian surfaces from the proposed framework. Fourth, an in-depth analysis on the merits of the learning-based reconstruction using EPIs is presented.

2. Related work

The main obstacle in light field imaging is the trade-off between spatial and angular resolution due to limited sensor resolution. Super-resolution techniques to improve spatial and angular resolution have been studied by many researchers [6, 7, 15, 36, 39]. In this paper, we mainly focus on approaches for improving the angular resolution of the light field.

2.1. Light field view synthesis

Zhang *et al.* [42] proposed a phase-based approach using a micro-baseline stereo pair. They applied a disparity (depth) assisted phase-based synthesis strategy to integrate the disparity information into the phase term when warping the input image to the novel view. However, their method was specifically designed for a micro-baseline stereo pair, and causes artifacts in the occluded regions when extrapolating novel views. Zhang *et al.* [41] described a patch-based approach for various light field editing tasks. In their work, the input depth map is decomposed into different depth layers and presented to the user to achieve the editing goals. However, these approaches rely heavily on quality of depth maps, which tends to fail in occluded, as well as glossy or specular, regions, thus often fail to produce promising results. In addition, they often focus on the quality of depth estimation, rather than the synthetic views themselves.

Boominathan *et al.* [7] capture the dense views from the hybrid light field. However, this would require lower spatial resolution. To break this trade-off, they use the high-resolution image in the center of the light field to provide the high-frequency detail information. Barnes *et al.* [5] proposed modern patch-based methods, which they efficiently scale to high-resolution photographs and collections of photos. So the data structure of their method can be used to accelerate the light field image super-resolution algorithm of Boominathan *et al.* [7]. For efficiency, they modify their algorithm by reducing patch feature descriptors to 20 dimensions with PCA.

Alternatively, some researches are based on sampling and consecutive reconstruction of the plenoptic function. For densely sampled light fields in which the disparity between the neighboring views does not exceed 1 pixel, novel views can be directly rendered by ray interpolation [23]. For sparsely sampled light fields, a reconstruction in Fourier domain has been investigated in some studies. Levin and Durand [22] proposed a linear view synthesis approach using a dimensionality gap light field prior to synthesize novel views from a set of images sampled with a circular pattern. Shi *et al.* [32] considered light field reconstruction as an optimization for sparsity in the continuous Fourier dimain. Their work sampled a small number of 1D viewpoint trajectories formed by a box and 2 diagonals to recover the full light field. However, these methods require the light field to be captured in a specific pattern, which limits its practical uses. Didyk *et al.* [11] used the phase information from a complex steerable pyramid decomposition to synthesize novel views with a small displacement; for large displacements, only low frequency components can be reconstructed.

2.2. Light field EPI structure

By taking advantage of EPI structure, Wanner and Goldluecke [36] employed structure tensor of an EPI to perform fast and robust local disparity estimation, then a TV-L^1 optimization scheme is applied to smooth the local result. Based on Wanner and Goldluecke's work, a certainty map was proposed to enforce visibility constrains on the initial estimated depth map in [24]. However, when implementing angular super-resolution, Wanner and Goldluecke [36] fell back into sub-aperture image space and warped the input images to synthesize novel views based on the disparity information. In contrast, Vagharshakyan *et al.* [34] considered the angular super-resolution as an inpainting problem on the EPI, and the angular aliasing could be suppressed in the Fourier domain. They therefore utilized an adapted discrete shearlet transform to reconstruct the light field from a sparse sampled light field. However, the reconstruction exhibited poor quality in the border regions, resulting in a reduction of angular extent. Moreover, the high frequency components in the EPI are also lost while using discrete shearlet to suppress high frequency leakage caused by angular aliasing.

2.3. Learning-based method

Recently, learning-based techniques have been explored for the light field reconstruction. Cho *et al.* [9] adopted a sparse-coding-based (SC) method to reconstruct light field using raw data. They generate image pairs using Barycentric interpolation. Yoon *et al.* [39] trained a deep neural network for spatial and angular super-resolution. However, the network used every two images to generate a novel view between them, thus it underused the potential of the full light field. Wang *et al.* [35] proposed several CNN architectures, one of which was developed for the EPI slices; however, the network is designed for material recognition, which is different with the EPI restoration task.

Also, some studies for maximizing the quality of synthetic views have been presented that are based on CNNs. Flynn *et al.* [14] proposed a deep learning method to synthesize novel views using a sequence of images with wide baselines. Kalantari *et al.* [20] used two sequential convolutional neural networks to model depth and color estimation simultaneously by minimizing the error between synthetic views and ground truth images. However, in that study, the network is trained using a fixed sampling pattern, which makes it unsuitable for universal applications. In addition, the approach

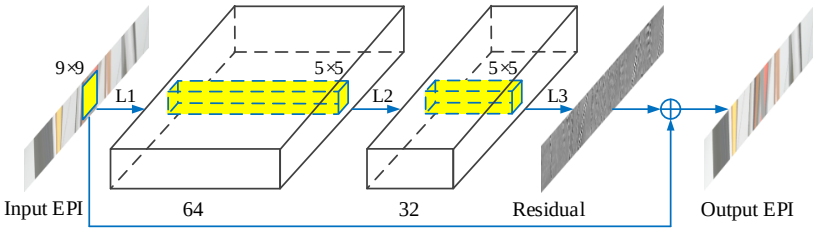

Figure 1: The detail restoration network is composed of three layers. The first and the second layers are followed by a rectified linear unit (ReLU). The final output of the network is the sum of the predicted residual (detail) and the input.

results in ghosting artifacts in the occluded regions and fails to handle some challenging cases.

3. CNN-based restoration in "blur-restoration-deblur" framework

3.1. CNN architecture

Based on Wu's method [38] and inspired by Dong *et al.* [12]'s network, the CNN architecture is shown in Fig. 1. The input EPI ($\mathbf{E}'_L$) is the low-resolution one, which is up-sampling by bicubic to the size as the high-resolution EPI. In the network, the desired output EPI $f(\mathbf{E}'_L)$ is the sum of the input $\mathbf{E}'_L$ and the predicted residual $\mathcal{R}(\mathbf{E}'_L)$:

$$f(\mathbf{E}'_L) = \mathbf{E}'_L + \mathcal{R}(\mathbf{E}'_L). \tag{1}$$

The CNN comprises three convolution layers. There are 64 filters of size $1 \times 9 \times 9$ in the first layer, where each filter operates on 9×9 spatial region across 64 channels (feature maps) and used for feature extraction. There are 32 filters of size $64 \times 5 \times 5$ in the second layer, and used for non-linear mapping. The last layer contains 1 filter of size $32 \times 5 \times 5$ used for detail reconstruction. Both the first and the second layers are followed by a rectified linear unit (ReLU).

3.2. Training detail

The desired residuals are $\mathbf{R} = \mathbf{E}' - \mathbf{E}'_L$, where $\mathbf{E}'$ are the blurred ground truth EPIs and $\mathbf{E}'_L$ are the blurred and interpolated low angular resolution EPIs. Our goal is to minimize the mean squared error $\frac{1}{2}\|\mathbf{E}' - f(\mathbf{E}'_L)\|^2$. However, due to the residual network we use, the loss function is now formulated as follows:

$$L = \frac{1}{n}\sum_{i=1}^{n} \|\mathbf{R}^{(i)} - \mathcal{R}(\mathbf{E}_L'^{(i)})\|^2, \tag{2}$$

where n is the number of training EPIs. The output of the network $\mathcal{R}(\mathbf{E}'_L)$ represents the restored detail, which must be added back to the input EPI $\mathbf{E}'_L$ to obtain the final high angular resolution EPI $f(\mathbf{E}'_L)$.

We use the Stanford Light Field Archive [3] (captured using a gantry system) as the training data. The blurred ground truth EPIs are decomposed to sub-EPIs of size 17×17 with stride 14. To avoid overfitting, we adopted data augmentation techniques [13, 21] that include flipping, downsampling the spatial resolution of the light field as well as adding Gaussian noise. To avoid the limitations of a fixed angular up-sampling factor, we use a scale augmentation technique. Specifically, we downsample some EPIs with a small angular extent by factor 4 and the desired output EPIs by factor 2, then upsample them to the original resolution. The network is trained by using the datasets downsampled by both factor 2 and factor 4. We use the cascade of the network for the EPIs that are required to be up-sampled by factor 4. In practice, we extract more than $8e6$ examples which is sufficient for the training. We select the mini-batches of size 64 as a trade-off between speed and convergence.

In the paper, we followed the conventional methods of image super-resolution to transform the EPIs into YCbCr space: only the Y channel (i.e., the luminance channel) is applied to the network. This is because the other two channels are blurrier than the Y channel and, thus, have less useful in the restoration [12].

To improve the convergence speed, we adjust the learning rate consistent with the increasing of the training iteration. The number of training iterations is 8×10^5 times. The learning rate is set to 0.01 initially and decreased by a factor of 10 every 0.25×10^5 iterations. When the training iterations are 5.0×10^5 , the learning rate is decreased to 0.0001 in two reduction steps. We initialize the filter weight of each layer using a Gaussian distribution with zero mean and standard deviation $1e^{-3}$. The momentum parameter is

set to 0.9. Training takes approximately 12 hours on GPU GTX 960 (Intel CPU E3-1231 running at 3.40GHz with 32GB of memory). The training model is implemented using the *Caffe* package [19].

4. Fourier analysis

In this section, we analyze the "blur-restoration-deblur" framework in the Fourier domain. Consider a simple scene composed of four points located at different depths. For an appropriate sampled light field, the resulting EPI contains four lines of different slopes, where each the disparity does not exceed 1 pixel (as shown in Fig. 2(a)). Fig. 2(b) shows the Fourier spectrum of the EPI, where the intersection angles with the Ω_u-axis are determined by the depths of the objects in the scene. In Fig. 2(b), we mark the Fourier spectrum of each line in Fig. 2(a) with arrow in its corresponding color.

We simulate the sparsely sampled light field in the angular domain by downsampling the light field in the angular dimensions, generating an angularly undersampled EPI whose disparity falls outside the one-pixel range. The sampling influences little on the point with a small disparity. However, for point with a large disparity, the sampling destroys high frequency detail in the angular dimension, producing copies of the Fourier spectrum, as shown in Fig. 2(c). Straightforward upsampling or CNN-based super-resolution will cause high frequency leakage from the other copies [18] (shown in Fig. 2(d)).

To overcome the aliasing in the EPI, Stewart *et al.* [33] applied a band-limit filter to reconstruct a light field in the Fourier domain. The filter preserves certain frequency component and simultaneously removes high frequency leakage by changing the shape of the filter (shown in the black dashed box in Fig. 2(d)). However, Liang and Ramamoorthi [25] indicated that the filter is depth-dependent, and simple reshaping the filter cannot reconstruct light field in all depth-of-field. Instead, in the proposed "blur-restoration-deblur" framework, we adopt a novel learning-based method to reconstruct sparsely sampled light field.

Specifically, we first balance the information between the spatial and angular information by a "blur" step. Unlike the band-limit filter described above, we use a simple 1D Gaussian kernel whose kernel size depends on the highest depth (disparity) of the light field. This "blur" step removes the high frequency components in the spatial dimension. Fig. 3(a) shows the Fourier spectrum of the blurred undersampled EPI. Compared with the Fourier spectrum of the undersampled EPI shown in Fig. 2(c), the copies lying in the high frequency regions are efficiently suppressed. The blurred EPI is upsampled to the desired angular resolution using bicubic interpolation.

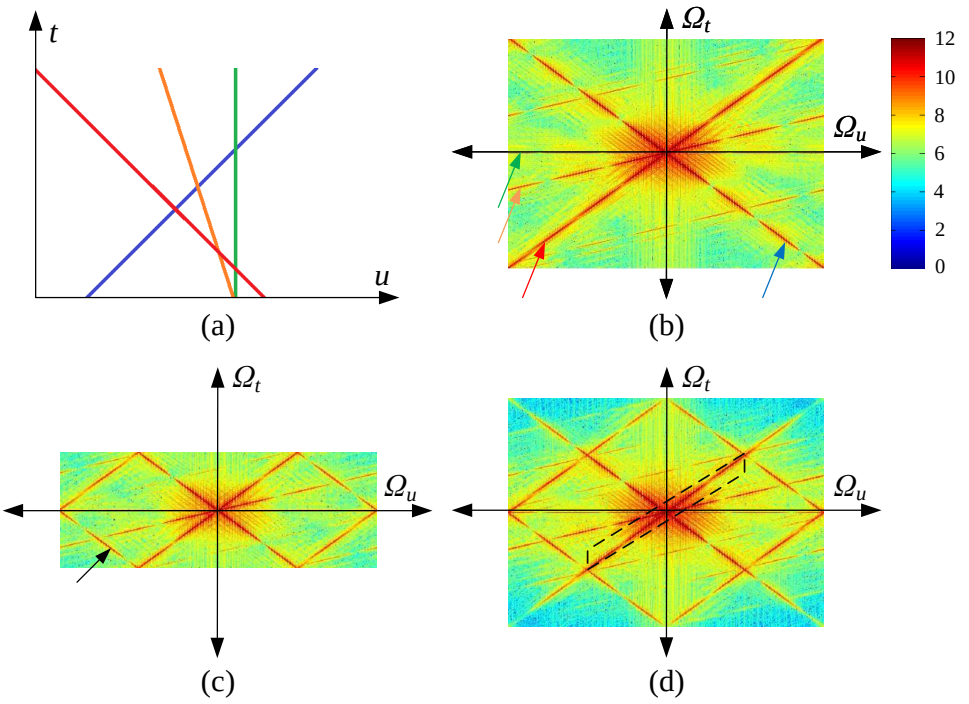

Figure 2: (a) An densely sampled EPI that contains four lines of different slopes. The disparity is no more than 1 pixel. (b) The Fourier spectrum of the EPI in (a), where the lines in (a) are marked with arrows in their corresponding colors. Note that the Fourier spectrum of the green line is occluded by the Ω_u-axis. (c) The light field is downsampled in the angular dimensions, producing an angularly undersampled EPI. The undersampling generates copies of the Fourier spectrum, where one of them is pointed out by a black arrow. (d) Directly CNN-based super-resolution causes high frequency leakage from the copies. A band-limit filter shown in the black dashed box can only reconstruct light field in a certain depth. The color bar on the right side of (b) shows the power range of the Fourier spectrum after taking the logarithm.

Then the CNN-based "restoration" step is performed to restore the angular detail. Fig. 3(b) shows the Fourier spectrum of the EPI after the "restoration" step. In the perspective of Fourier spectrum, the CNN is trained to restore the high frequency components rather than the high frequency copies that lead to aliasing.

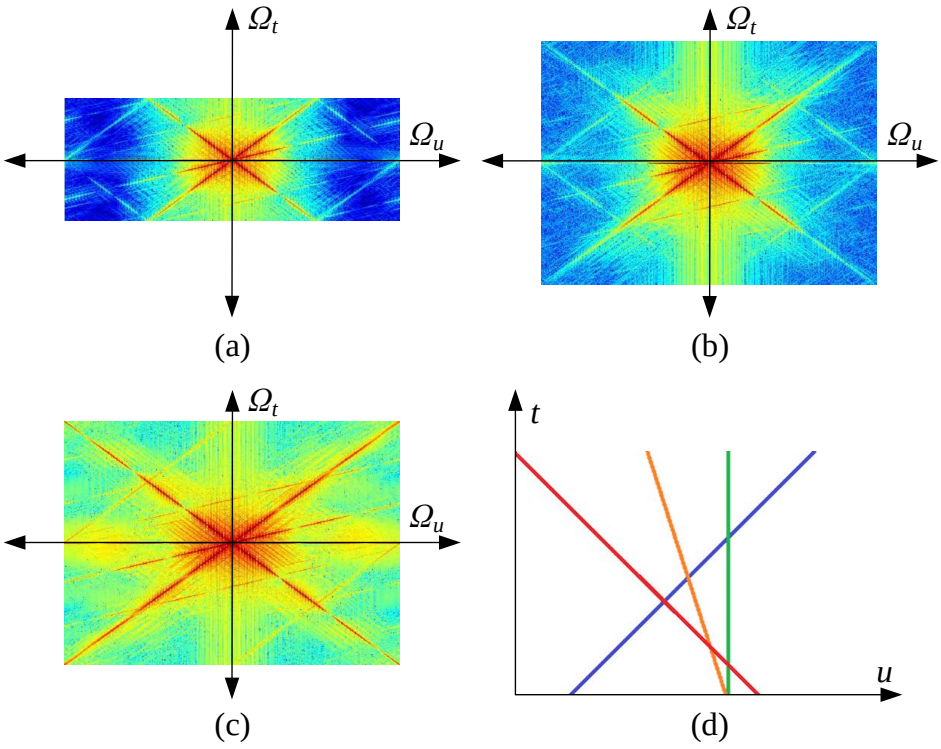

Figure 3: (a) The Fourier spectrum of the undersampled EPI after the "blur" step, where the high frequency copies are efficiently suppressed. (b) The Fourier spectrum of the EPI after the "restoration" step. Compared with the Fourier spectrum in (a), the high frequency components are restored, while the high copies that lead to aliasing are remained unchanged. (c) The Fourier spectrum of the EPI after being processed by the entire framework. (d) The superresolved EPI produced by the proposed framework. The color bar on the right side of Fig. 2(b) shows the power range of the Fourier spectrum after taking the logarithm.

The "deblur" step using the selected Gaussian kernel is adopted to recover the high frequency components in the spatial dimension, which is an inverse operation with respect to the "blur" step. Fig. 3(c) and (d) shows the Fourier spectrum and the EPI, respectively, after being processed by the entire "blur-restoration-deblur" framework. Due to the removing of the high

frequency copies in the Fourier spectrum, the EPI is finally super-resolved without aliasing.

5. Depth assisted rendering

The proposed framework by Wu *et al.* still has its limitations. They use EPI blur to extract the low frequency components of the EPI in the spatial dimension, where the size of the blur kernel is determined by the largest disparity between the input neighboring views. The non-blind deblur is not able to recover high quality EPIs when the kernel size is too large, and the maximum disparity we can handle when using the proposed "blur-restoration-deblur" framework is 5 pixels. In our strategy, we exploit depth information to handle large disparity data, such as multi-view stereo data. In addition, at least 3 views should be used in each angular dimension to provide enough information for the bicubic interpolation.

So we extend this approach for rendering novel views using multi-view input (e.g., multi-view stereo data) and a depth (disparity) map. Unlike the existing depth image-based rendering (DIBR) techniques that are extremely sensitive to the disparity quality, only a raw disparity map is required for the proposed rendering approach. In addition, the novel view rendering method inherits the capacity of generating plausible result in occlusion regions, and non-Lambertian surfaces from the proposed learning-based framework.

5.1. EPI shearing

First, we take the 2D EPI for example, x represents the spatial domain, and u represents the angular domain. So the EPI can be rewrite as,

$$L_\alpha(x,u) = L_0\left(x + u\left(1 - \frac{1}{\alpha}\right), u\right), \tag{3}$$

L_0 denotes the original EPI, L_α denotes the sheared EPI by a value of α.

Then, the extended 4D form can be expressed as, which is similar to the equation from Ng *et al.*, [28]

$$L_\alpha(x,y,u,v) = L_0\left(x + u\left(1 - \frac{1}{\alpha}\right), y + v\left(1 - \frac{1}{\alpha}\right), u, v\right), \tag{4}$$

where, x, y represent the spatial domain, and u, v represent the angular domain. Eq.4 is the formula which establishes the relationship between the

Figure 4: The comparison between original EPIs with their sheared versions.

original EPIs and thier sheared versions. Fig. 4 shows the effect before and after the shearing.

5.2. Implementation

The main obstacle for reconstructing an EPI with large disparity is the information asymmetry, instead of applying a blur kernel to mitigate the asymmetry (which actually fail to work in this circumstance), we first shear the EPI to an appropriate disparity range with the assistance of a disparity map, then the novel view rendering can be considered as the super-resolving problem on the sheared EPI using the proposed "blur-restoration-deblur" framework.

Specifically, consider an EPI E_L (where L denotes low angular resolution) and its discretized disparity D (see Fig. 5(a)), the collection of shear values is equal to the collection of discretized disparity values in D, we first shear the EPI with each shear value, constructing a set of sheared EPIs $\{S(E_L^{D_1}), S(E_L^{D_2}), ..., S(E_L^{D_N})\}$ (see Fig. 5(b)), where S denotes the shear operation and N is the number of the discretized disparity values. Then the proposed "blur-restoration-deblur" framework is applied to super-resolve the sheared EPIs in the angular dimension (see Fig. 5(c)), generating a set of high angular resolution EPIs $\{S(E_H^{D_1}), S(E_H^{D_2}), ..., S(E_H^{D_N})\}$. For regions sheared using their corresponding disparity (demonstrated as the brighter regions in Fig. 5(b) and (c)), we can obtain super-resolved results with best performance; however, other regions (demonstrated as the darker regions in Fig. 5(b) and (c)) are sheared with improper shear value, causing aliasing effects in the corresponding regions of the super-resolved EPIs. Fig. 5(e) shows a close-up of one of those super-resolved regions. To combine all the best super-resolved regions and obtain the final EPI E_H, we implement the following steps: the super-resolved EPIs $\{S(E_H^{D_1}), S(E_H^{D_2}), ..., S(E_H^{D_N})\}$ are first

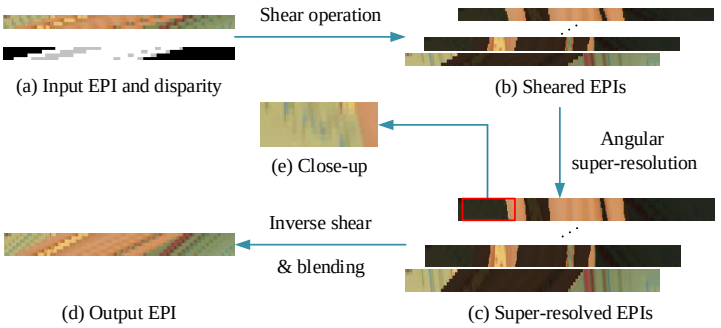

Figure 5: Pipeline of the proposed depth assisted rendering approach using the "blur-restoration-deblur" framework. (a) The input EPI and its disparity. The disparity is discretized for the shear operation; (b) The input EPI is sheared by each shear value, constructing a set of sheared EPIs. The brighter regions are sheared using their corresponding disparity, but the darker regions are not; (c) The super-resolved EPIs using the proposed "blur-restoration-deblur" framework; (d) The final high angular resolution EPI is obtained by using the inverse shear and blending operation; (e) Close-up of one of the super-resolved EPIs in (c).

inversely sheared using shear value $\{\frac{D_1}{\alpha}, \frac{D_2}{\alpha}, ..., \frac{D_N}{\alpha}\}$, where α is the super-resolution factor. Then the processed EPIs, denoted as $\{E_H^{D_1}, E_H^{D_2}, ..., E_H^{D_N}\}$, are blended using the following equation:

$$E_H = \sum_{i=1}^{N} E_H^{D_i} W_i, \tag{5}$$

and the weight W_i is equal to 1 for regions sheared using the corresponding disparity, and 0 for other regions. Fig. 5(d) shows the final high angular resolution EPI E_H.

5.3. Results

We evaluate the proposed application for depth assisted rendering on Middleburry Stereo Datasets (including 2005 datasets and 2006 datasets [16, 30, 31]), which contains multiple views with large disparities (compared with light field data). We employ the CostFilter [29] for disparity estimation using only two views in the input data. The previous DIBR technique used

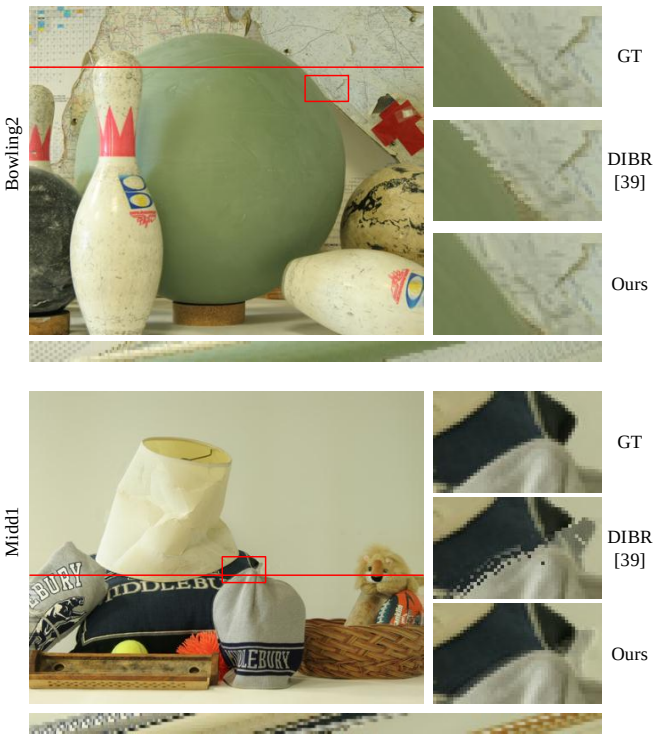

Figure 6: Depth assisted rendering results on Middleburry datasets. The bottom shows the EPI super-resolved by the proposed approach.

	Baby1	*Books*	*Bowling2*	*Reindeer*	*Moebius*
DIBR [8]	38.10/0.9907	35.95/0.9958	35.92/0.9947	31.50/0.9886	35.97/0.9947
Ours	**41.07/0.9944**	**37.34/0.9959**	**38.22/0.9963**	**33.80/0.9906**	**38.41/0.9964**

Table 1: Quantitative results (PSNR / SSIM) of the rendered views (results are averaged on the three novel views).

for angular super-resolution of light fields [8] is applied for the comparison. Each scene in the datasets contains 7 views, where 4 of them are used as input and the others are for comparison.

Fig. 6 shows the visual comparison of rendered novel views against the ground truth, and Table 1 offers the relevant numerical results in terms of PSNR and SSIM. As we can see from the figure, results by DIBR method [8]

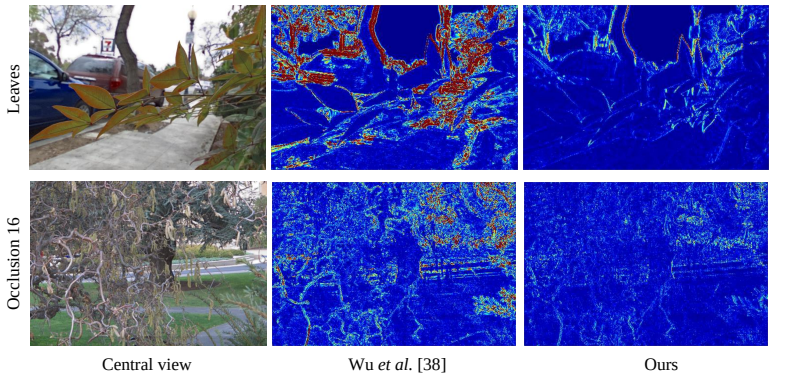

Figure 7: The error map comparison between Wu *et al.*'s [38] method and our method.

usually contain ghosting artifacts in occlusion boundaries (see the close-up in Fig. 6). In addition, due to extensive textureless regions in the last two cases (*Midd1* and *Monopoly*), the stereo matching approach being employed has failed to produce plausible disparity maps, which causes great influence on results by DIBR method [8]. However, the proposed approach is able to render high quality views in both cases.

In spite that the proposed approach exploits depth information to render novel views, the "blur-restoration-deblur" framework can tolerate disparity error within a certain range. This property makes the proposed depth assisted approach more robust to common disparity noise in occlusion and textureless regions.

Finally, since our method is based on Wu *et al.*'s [38] work, the results of our method should be compared with their method. We take the *Leaves* data [20] and *Occlusions 16* [3] data as the examples. In order to test cases in the large disparity (typically exceed 4 pixels), we take samples from every three viewpoints to expand the disparity. The *Leaves* case includes complex structure and the *Occlusions 16* contains complicated occlusions that are challenging. Both of them have the large disparity. Fig. 7 shows some examples that the error map results between the groundtruth and super-resolved image in angular domain. The results of Wu *et al.*'s method are quite blurry around the occluded regions such as branched and leaves. As demonstrated in the error maps of the results, the proposed approach achieves a high performance.

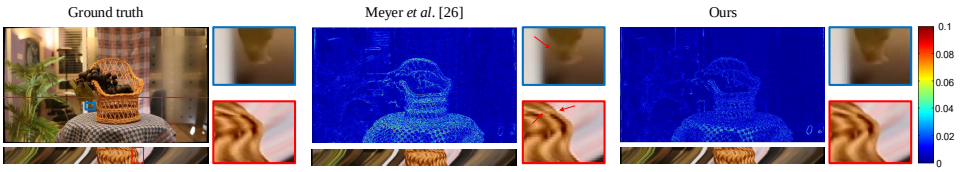

Figure 8: Results of the super-resolution of unstructured light field on the *Basket* case from the dataset provided by Yücer *et al.* [40].

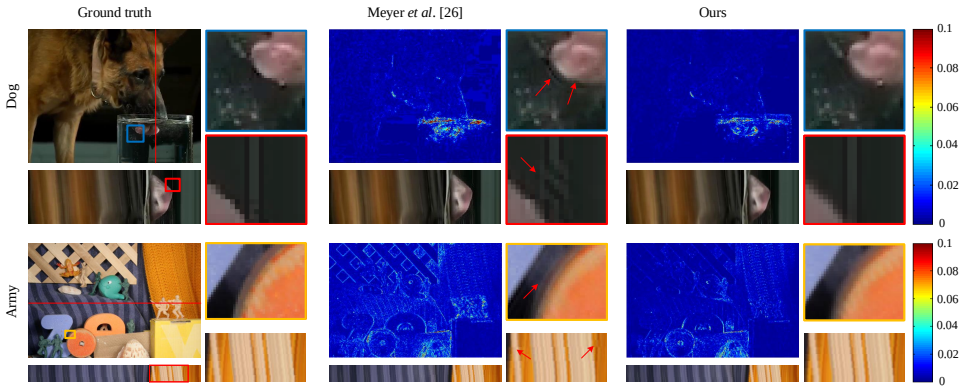

Figure 9: Comparison of frame interpolation results for video sequences. The interpolation results of a high frame rate video footage are shown in the top of the figure. The red boxes show the close-up version of the 2D slices, where the result by Meyer *et al.* [26] appears ghosting effects. The interpolation results on the *Army* case from Middlebury dataset [4] are shown in the bottom of the figure. The result produced by Meyer *et al.* [26] shows ghosting effects and structure discontinuity (see the close-up version of the 2D slice).

6. Applications

We implement two different applications based on the proposed "blur-restoration-deblur" framework: angular super-resolution for unstructured light field and video interpolation.

The proposed approach is a depth-free framework that restores the angular detail based on CNN without the need of geometry calibration and depth estimation. Therefore, our framework is potentially capable to handle

unstructured input, such as an unstructured light field and a video sequence. This application can improve the accuracy of reconstructed 3D model [40] when applied to the interpolation of unstructured light field, and further used for frame rate conversion, slow motion video and motion interpolation [26] when applied to video interpolation.

6.1. Angular super-resolution for unstructured light field.

Unlike common light field that is obtained by carefully calibrated camera(s), an unstructured light field [10] is captured, e.g., by a hand-held commodity camera. Therefore, ordinary angular super-resolution methods that synthesize novel views based on depth estimation usually fail to yield reasonable results. On the contrary, the proposed framework implicitly reconstruct a light field by restoring the angular detail on EPI, and thus, has the capacity to super-resolve unstructured light field.

We demonstrate this application using the *Basket* case from the dataset provided by Yücer *et al.* [40] (see Fig. 8). The original light field contains 49 views, in which 25 views is used as input for the super-resolution. We compare our method against the approach by Meyer *et al.* [26]. They introduces structure discontinuity when comparing the resulting 2D slices (see the close-up version in Fig. 8). The averaged PSNR values are 41.60 for the proposed framework and 40.43 for the approach by Meyer *et al.* [26].

6.2. Video interpolation

For the frame interpolation of a video sequence, we first obtain 2D slices, which is similar with the EPIs in a light field, by gathering horizontal (or vertical) lines at a certain coordinate in each frame; then the proposed framework is implemented on them. We demonstrate the interpolation result using a high frame rate video footage. The original video footage contains 97 frames, and only 25 frames are used as the input. We compare the proposed framework against phase-based frame interpolation approach proposed by Meyer *et al.* [26]. The interpolation results of the 78^{th} frame are shown in the top of Fig. 9. The phase-based frame interpolation shows ghosting effects when compared the 2D slice (see the close-up version in the red box). The quantitative results in terms of PSNR (averaged on the interpolated frames) are 47.89 for the proposed framework and 45.56 for the approach by Meyer *et al.* [26].

The bottom of Fig. 9 shows the interpolation results on the *Army* case from Middlebury optical flow dataset [4]. We apply 7 frames in the sequence,

in which 4 frames are applied as the input. The sequence contains multiple moving directions, which is challenge. The figure shows the interpolation results of the 2nd frame. The result produced by the proposed framework introduces slightly blurry in some large motion parts, such as the wheel shown in the yellow box; while the result produced by Meyer *et al.*'s approach [26] contains less detail and introduces ghosting effects. In addition, the phase-based interpolation approach [26] uses only two frames to produce intermediate images, introducing structure discontinuity in the 2D slices. The averaged PSNR values are 41.76 for the proposed framework and 39.31 for the approach by Meyer *et al.* [26].

7. Conclusion and discussion

In this paper, based on Wu *et al.* [38], we further study extended applications including depth enhancement using reconstructed high angular resolution light field, interpolation for unstructured input and depth assisted rendering. In the following, we provide an in-depth analysis on the merits of the learning-based reconstruction using EPIs as well as the limitation of the proposed framework that should be overcome in the future work.

References

[1] Lytro, `https://www.lytro.com/`.

[2] RayTrix, 3D light field camera technology, `http://www.raytrix.de/`.

[3] Stanford (New) Light Field Archive, `http://lightfield.stanford.edu/lfs.html`.

[4] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, *A database and evaluation methodology for optical flow*, International Journal of Computer Vision **92** (2011), no. 1, 1–31.

[5] C. Barnes, F. L. Zhang, L. Lou, X. Wu, and S. M. Hu, *Patchtable: efficient patch queries for large datasets and applications*, Acm Transactions on Graphics **34** (2015), no. 4, 97.

[6] T. E. Bishop and P. Favaro, *The light field camera: Extended depth of field, aliasing, and superresolution*, IEEE TPAMI **34** (2012), no. 5, 972–986.

[7] V. Boominathan, K. Mitra, and A. Veeraraghavan, *Improving resolution and depth-of-field of light field cameras using a hybrid imaging system*, in: ICCP, pages 1–10. IEEE, 2014.

[8] G. Chaurasia, S. Duchêne, O. Sorkine-Hornung, and G. Drettakis, *Depth synthesis and local warps for plausible image-based navigation*, ACM TOG **32** (2013).

[9] D. Cho, M. Lee, S. Kim, and Y.-W. Tai, *Modeling the calibration pipeline of the lytro camera for high quality light-field image reconstruction*, in: ICCV (2013).

[10] A. Davis, M. Levoy, and F. Durand, *Unstructured light fields*, Comput. Graph. Forum **31** (2012), no. 2.

[11] P. Didyk, P. Sitthi-Amorn, W. Freeman, F. Durand, and W. Matusik, *Joint view expansion and filtering for automultiscopic 3d displays*, in: Siggraph, pages 3906–3913, 2015.

[12] C. Dong, C. C. Loy, K. He, and X. Tang, *Learning a deep convolutional network for image super-resolution*, in: ECCV, pages 184–199. Springer, 2014.

[13] D. Eigen, C. Puhrsch, and R. Fergus, *Depth map prediction from a single image using a multi-scale deep network*, in: Advances in Neural Information Processing Systems, pages 2366–2374, 2014.

[14] J. Flynn, I. Neulander, J. Philbin, and N. Snavely, *Deepstereo: Learning to predict new views from the world's imagery*, in: CVPR, 2015.

[15] X. Guo, Z. Yu, S. B. Kang, H. Lin, and J. Yu, *Enhancing light fields through ray-space stitching*, IEEE TVCG **22** (2016), no. 7, 1852–1861.

[16] H. Hirschmüller and D. Scharstein, *Evaluation of cost functions for stereo matching*, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2007.

[17] I. Ihrke, J. F. Restrepo, and L. Mignard-Debise, *Principles of light field imaging: Briefly revisiting 25 years of research*, IEEE Signal Process. Mag. **33** (2016), no. 5, 59–69.

[18] A. Isaksen, L. McMillan, and S. J. Gortler, *Dynamically reparameterized light fields*, in: Proceedings of the 27th annual conference on Computer graphics and interactive techniques, pages 297–306. ACM Press/Addison-Wesley Publishing Co., 2000.

[19] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, *Caffe: Convolutional architecture for fast feature embedding*, in: ACM MM, pages 675–678. ACM, 2014.

[20] N. K. Kalantari, T.-C. Wang, and R. Ramamoorthi, *Learning-based view synthesis for light field cameras*, ACM Transactions on Graphics (TOG) **35** (2016), no. 6.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *Imagenet classification with deep convolutional neural networks*, in: Advances in Neural Information Processing Systems, pages 1097–1105, 2012.

[22] A. Levin and F. Durand, *Linear view synthesis using a dimensionality gap light field prior*, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010.

[23] M. Levoy and P. Hanrahan, *Light field rendering*, in: Siggraph, pages 31–42, 1996.

[24] J. Li, M. Lu, and Z.-N. Li, *Continuous depth map reconstruction from light fields*, IEEE TIP **24** (2015), no. 11, 3257–3265.

[25] C.-K. Liang and R. Ramamoorthi, *A light transport framework for lenslet light field cameras*, in: Siggraph, volume 34, pages 1–19, 2015.

[26] S. Meyer, O. Wang, H. Zimmer, M. Grosse, and A. Sorkine-Hornung, *Phase-based frame interpolation for video*, in: CVPR, pages 1410–1418, 2015.

[27] S. Pujades, F. Devernay, and B. Goldluecke, *Bayesian view synthesis and image-based rendering principles*, in: CVPR, pages 3906–3913, 2014.

[28] N. Ren, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, *Light field photography with a hand-held plenopic camera*, CSTR 2005-02, 2005.

[29] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, *Fast costvolume filtering for visual correspondence and beyond*, in: IEEE TPAMI, volume 35, pages 504–511, 2011.

[30] D. Scharstein and C. Pal, *Learning conditional random fields for stereo*, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2007.

[31] D. Scharstein and R. Szeliski, *High-accuracy stereo depth maps using structured light*, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 1, pages 195–202, 2003.

[32] L. Shi, H. Hassanieh, A. Davis, D. Katabi, and F. Durand, *Light field reconstruction using sparsity in the continuous fourier domain*, ACM TOG **34** (2014), no. 1, 12.

[33] J. Stewart, J. Yu, S. J. Gortler, and L. McMillan, *A new reconstruction filter for undersampled light fields*, in: Eurographics Workshop on Rendering, pages 150–156, 2003.

[34] S. Vagharshakyan, R. Bregovic, and A. Gotchev, *Image based rendering technique via sparse representation in shearlet domain*, in: ICIP, pages 1379–1383. IEEE, 2015.

[35] T.-C. Wang, J.-Y. Zhu, E. Hiroaki, M. Chandraker, A. A. Efros, and R. Ramamoorthi, *A 4d light-field dataset and cnn architectures for material recognition*, in: ECCV, pages 121–138. Springer, 2016.

[36] S. Wanner and B. Goldluecke, *Variational light field analysis for disparity estimation and super-resolution*, IEEE TPAMI **36** (2014), no. 3, 606–619.

[37] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, *High performance imaging using large camera arrays*, in: ACM Transactions on Graphics (TOG), volume 24, pages 765–776. ACM, 2005.

[38] M. Zhao, G. Wu, Y. Liu, and X. Hao, *How depth estimation in light fields can benefit from super-resolution?*, International Journal of Advanced Robotic Systems, volume 15, no. 1, pages 1729881417748446. 2018.

[39] Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, and I. So Kweon, *Learning a deep convolutional network for light-field image super-resolution*, in: CVPRW, pages 24–32, 2015.

[40] K. Yücer, A. Sorkine-Hornung, O. Wang, and O. Sorkine-Hornung, *Efficient 3D object segmentation from densely sampled light fields with applications to 3D reconstruction*, ACM Transactions on Graphics **35** (2016), no. 3.

[41] F.-L. Zhang, J. Wang, E. Shechtman, Z.-Y. Zhou, J.-X. Shi, and S.-M. Hu, *Plenopatch: Patch-based plenoptic image manipulation*, IEEE TVCG.

[42] Z. Zhang, Y. Liu, and Q. Dai, *Light field from micro-baseline image pair*, in: CVPR, pages 3800–3809, 2015.

Zhengzhou Institute of Surveying and Mapping
Zhengzhou, Henan 450001, China
E-mail address: mandanzhao@163.com

Zhengzhou Institute of Surveying and Mapping
Zhengzhou, Henan 450001, China
E-mail address: xiangyanghao2004@163.com