Implicit Epipolar Geometric Function based Light Field Continuous Angular Representation

Lin Zhong  Bangcheng Zong*  Qiming Wang  Junle Yu  Wenhui Zhou†
Hangzhou Dianzi University, Hangzhou, China
{zhonglin, bc_zhong, cheamy, yjl, zhouwenhui}@hdu.edu.cn

Abstract

Light field plays an important role in many different applications such as virtual reality, microscopy and computational photography. However, low angular resolution limits the further application of light field. The existing state of the art light field angular super-resolution reconstruction methods can only achieve limited fixed-scale angular super-resolution. This paper focuses on a continuous arbitrary-scale light field angular super-resolution via introducing the implicit neural representation into the light field two-plane parametrization. Specifically, we first formulate a 4D implicit epipolar geometric function for light field continuous angular representation. Considering it is difficult and inefficient to directly learn this 4D implicit function, a divide-and-conquer learning strategy and a spatial information embedded encoder are then proposed to convert the 4D implicit function learning into a joint learning of 2D local implicit functions. Furthermore, we design a special epipolar geometric convolution block (EPIBlock) to encode the light field epipolar constraint information. Experiments on both synthetic and real-world light field datasets demonstrate that our method exhibits not only significant superiority in fixed-scale angular super-resolution, but also achieves arbitrary high magnification light field super-resolution while still maintaining the clear light field epipolar geometric structure.

1. Introduction

Light field imaging [12, 16, 19], recording 4D spatial-angular information of incident light rays, is always an important research hotspot in the field of computational imaging [32]. Compared to traditional 2D imaging, the additional angular information in the light field contains more helpful multi-view cues for scene analysis and understanding, and the angular resolution is also a key quality-control factor for light field display. However, there is a tread-off between light field spatial and angular resolution. In recent years, light field angular super-resolution reconstruction (SR), as an alternative method, has attracted significant research interest [31].

Although numerous state of the art light field super-resolution methods has been proposed [1, 8, 11, 13, 28, 31, 35], they can only reconstruct a sparsely sampled light field to a fixed angular resolution (up to $7 \times 7$ or $9 \times 9$), and cannot achieve continuous arbitrary-scale resolution reconstruction because the complexity of their methods is coupled with a fixed resolution.

The most recently proposed implicit neural representation [3, 18, 36] provides a creative and innovative idea for image super-resolution and view synthesis, via using neural learning to parameterize an implicitly defined, continuous, differentiable 2D/3D signal (image or scene). In this paper, we introduce the implicit neural representation into the 4D light field two-plane parameterization, and design a 4D light field implicit epipolar geometric function for light field continuous angular representation. Obviously, it is difficult and inefficient to directly learn this 4D implicit function,
especially for limited light field data. Herein, following the divide-and-conquer strategy, we decouple the horizontal and vertical angular representation by fixing an angular and spatial coordinate of the 4D light field, and decompose the 4D implicit epipolar geometric function into a set of 2D local implicit horizontal/vertical angular functions. Then we convert the 4D implicit function learning into a joint learning of 2D local implicit functions. Moreover, to preserve spatial consistency among local implicit functions, a spatial information embedded encoder is designed to embed the spatial consistency into the local latent angular representations, a spatial information embedded encoder is designed to preserve spatial consistency among local implicit functions.

In summary, our main contributions include:

- An implicit epipolar geometric function-based light field continuous angular representation, which can reconstruct arbitrary angular resolution light field with clear epipolar geometric structure.
- A divide-and-conquer learning strategy for 4D light field implicit epipolar geometric function, which converts the 4D implicit function learning into joint learning of 2D local implicit functions.
- An EPIBlock is designed to encode the spatial consistency of adjacent epipolar plane images (EPIs) into the local implicit function learning, which alleviates the limitation of lacking spatial structure information and constraints only using a single EPI.

2. Related Work

2.1. Light Field Super-resolution

Light field super-resolution reconstruction aims to enhance the spatial and/or angular dimensions of 4D light field images. Here we mainly focus on the light field angular super-resolution reconstruction, which are generally divided into the sub-aperture image (SAI) based and the EPI based methods.

**SAI-based Methods** The state of the art SAI-based methods usually adopt deep learning to extract spatial and angular semantic features from light field SAIIs, and then synthesize virtual views at specified angular coordinates. Yoon et al. [35] first proposed a deep convolutional neural network (CNN) for light field spatial and angular super-resolution reconstruction. LFNet [26] incorporated an implicitly multi-scale fusion scheme into a bidirectional recurrent convolutional neural network to accumulate contextual information for light field super-resolution. Gul et al. [8] took raw light field data as input, and used a CNN to enhance the spatial and angular resolution. Yeung et al. [33] designed a spatial-angular separable CNN for light field spatial super-resolution. Ko et al. [15] presented an adaptive feature remixing approach for spatial and angular super-resolution. Wang et al. [24] proposed a class of domain-specific convolutions to disentangle the spatial and angular coupling information and developed three network, DistgSSR, DistgASR and DistgDisp, for spatial super-resolution, angular super-resolution and disparity estimation, respectively.

**EPI based Methods** Inspired by 2D image super-resolution, Wu et al. [31,32] contributed a “blur-restoration-deblur” framework to recover high frequency details of multiple EPIs, which can alleviate the the problem of asymmetry information between the spatial and angular dimensions caused by sparse angular sampling. They also proposed a learning-based light field reconstruction approach by fusing a set of sheared EPIs [30]. Wang et al. [25] proposed a multi-EPI based approach, which applied a 3D convolution layer to recover details on horizontal and vertical EPIs in turn. They further improved their method by adding the EPI structural recovery loss function [27].

2.2. Implicit neural representation

Recent studies have demonstrated that implicit parameterization of continuous functions using a trained multilayer perceptron (MLP) is an efficient alternative to traditional convolution. Mildenhall et al. [18] used an MLP neural network to implicitly learn a static 3D scene. Yu et al. [36] took the pixel-aligned spatial image features as input, which allows the framework to train and learn scene priors from a set of multi-view images, and then synthesize views from one or several input images. Li et al. [17] predicted a 4-channel image (RGB and volume density) at arbitrary depth values to jointly reconstruct the camera frustum and fill in occluded contents from a single image. Peng et al. [20] proposed a novel human body representation that assumes that the learned neural representations at different frames share the same set of latent codes anchored to a deformable mesh. Wang et al. [34] developed a hybrid neural surface representation to impose geometry-aware sampling and regularization, which can significantly improve the fidelity of reconstructions. Chen et al. [4] performed the generation of 2D shapes for simple numbers from the latent space. They further proposed a local implicit image function (LIIF), which takes image coordinates and 2D deep features as input, and predicts the RGB values at a given coordinates [3]. Sitzmann et al. [23] used MLPs with periodic activation functions for implicit neural representation instead of ReLU-MLPs, and demonstrated that it can model natural images with higher quality.

3. Method

3.1. Light field EPI representation

Light field EPI contains rich epipolar geometry information of light field, such as the linear structure formed by the
intersection of the epipolar plane and the camera plane. An EPI can be obtained by stacking pixels from a same row (or column) of a row (or column) of SAI.

Specifically, as for a two-plane parameterized light field $I(u, v, s, t)$ with resolution of $U \times V \times W \times H$, a horizontal EPI $I_{epi}(u, t) (u, s)$ can be obtained by stacking the $t_i$-th row pixels of the $v_i$-th row SAI, and its horizontal neighborhood EPI $I_{neib}(u, t_i) (u, s)$ is built by stacking $2m + 1$ adjacent horizontal EPIs together along the $u$ axis, including $I_{epi}(u, t_i-a) (u, s), \ldots, I_{epi}(u, t_i-1) (u, s), I_{epi}(u, t_i) (u, s), I_{epi}(u, t_i+1) (u, s), \ldots, I_{epi}(u, t_i+a) (u, s)$. The resolution of a horizontal neighborhood EPI is $(2m + 1)U \times W$. Similarly, a vertical EPI is defined as $I_{epi}(u, v) (v, t)$, and its vertical neighborhood EPI $I_{neib}(u, v) (v, t)$ can be built by stacking $2m + 1$ adjacent vertical EPIs together along the $v$ axis, including $I_{epi}(u-a, v) (v, t), \ldots, I_{epi}(u-1, v) (v, t), I_{epi}(u, v) (v, t), I_{epi}(u+1, v) (v, t), \ldots, I_{epi}(u+a, v) (v, t)$. The resolution of a vertical neighborhood EPI is $(2m + 1)V \times H$. In our experiments, the $m$ is set to 2 by default, as shown in Figure 2.

We construct a light field EPI tuple with a light field EPI and its neighborhood EPI. For simplicity, considering $U = V$ and $W = H$ in most of synthetic light field datasets, we uniformly denote the light field EPI (whether horizontal or vertical ones) as $I_{epi, basis}$ and the neighborhood EPI (whether horizontal or vertical neighbors) as $I_{epi, neib}$. Then we combine each EPI $I_{epi, basis}$ and its corresponding neighborhood EPI $I_{epi, neib}$ into a light field EPI tuple $I_{epi, group} = \{I_{epi, basis}, I_{epi, neib}\}$.

### 3.2. Overview of our network

In this paper, we propose an implicit epipolar geometric function based light field angular super-resolution (named as “LFEIFASR”). It mainly includes a spatial-angular fusion module, an epipolar geometry residuals dense network, and an epipolar geometry implicit function learning, as shown in Figure 3.

### 3.3. Spatial-angular fusion module

Although a light field EPI implies rich angular information, it lacks spatial information because of itself construction limitation. To address this issue, we supplement the spatial constraint information by building light field EPI tuples during the generation of training set. We proposed two specific convolution to extract angular and spatial information from each EPI $I_{epi, basis}$ and its neighborhood EPI array $I_{epi, neib}$, respectively.

For a LR (low resolution) EPI $I_{epi, basis}$, we apply a convolution layer with a $3 \times 9$ convolution kernel, named as “Angular Conv”, to extract the angular continuous features $H_{Angular, Conv}$ from $I_{epi, basis}$.

For a LR neighborhood EPI $I_{epi, neib}$, we design an expanded convolution, named as “Spatial Conv”, to extract the spatial continuous features $H_{Spatial, Conv}$ from pixels of the same row in adjacent EPIs, by setting the convolution kernel size to be $(2m + 1, 2m + 1)$ and the expansion coefficient to be $(2m + 1, 1)$.

By summing up the angular and spatial continuous information, we obtain the spatial-angular fusion information $F_{mixed}$ as follows.

$$F_{mixed} = H_{Spatial, Conv}(I_{epi, neib}) + H_{Angular, Conv}(I_{epi, basis})$$ (1)

### 3.4. Epipolar geometry residuals dense network

Traditional convolution is only suitable for extracting the features within the local receptive field, and it is difficult to effectively cover all pixels in EPI linear structures with various slopes. To alleviate this limitation, inspired by the skip connection structures of ResNet [9] and DenseNet [37], we propose an Epipolar Geometric Convolution Block (named as “EPIBlock”) and an epipolar geometric convolution (named as “EPIConv”) to extract epipolar geometric features from light field EPI tuples, as shown in Figure 3.

Each EPIBlock contains $C$ EPIConvs, and the structure of EPIConv is shown in Figure 4. Specifically, the output from the $U \times 1$ convolution is reshaped to the same size of the feature produced by the $1 \times 7$ convolution. Then the extracted feature is summed and fused by a $3 \times 3$ convolution.
Our epipolar geometry residuals dense network consists of \( D \) EPIBlocks, and each EPIBlock comprising of \( C \) EPICons, which forms a DenseNet structure to fully extract epipolar geometric features, that is each EPIConv accepts feature information from all previous EPICons. Therefore, our network leverages global residual connections to extract deep feature information in the EPIs, and to fuse the outputs of the EPIBlocks by concatenation operation. In our experiments, the channel numbers of all convolutional layers in EPICons are set to be 64, as shown in Eqs (7)-(8):

\[
F_d = H_{EPI\text{Block},d}(F_{d-1}) = H_{EPI\text{Block},d}(H_{EPI\text{Block},d-1}(\cdots(H_{EPI\text{Block},1}(F_0))))
\]  

\[
F_{\text{epi,global}} = H_{\text{TailConv}}(F_1, F_0, F_1, \ldots, F_D)
\]

where \( F_d \) denotes the epipolar geometry feature produced by the \((d-1)\)th EPIBlock. The epipolar geometry global feature denoted as \( F_{\text{epi,global}} \) in the EPIs can be extracted by applying a \( 3 \times 3 \) convolution (denoted as \( \text{TailConv} \)) to a tensor concatenating output from each EPIBlock, the \( F_{\text{epi,global}} \) has the same dimensions as the EPI image with low angular resolution.
3.5. Epipolar geometric implicit function learning

Traditional super-resolution methods primarily used 2D interpolation based image upsampling [5, 6, 14, 21], which usually recover the missing RGB pixels based on explicit information from surrounding pixels, by learning the discrete representation of 2D images. Herein, we follow LIIF [3] to use an epipolar geometric implicit function learning strategy.

Firstly, we use the coordinate index matrix generation method to get the low-resolution (LR) coordinate map (i.e., CoordMapLR) with LR epipolar geometry implicit code (i.e., Fepl,global), and then the high angular resolution (HR) sampling coordinate map (i.e., CoordMapHR-Sampling) is obtained from CoordMapLR through the nearest neighbor interpolation. The HR coordinate (HR Coord) is used to generate high angular resolution coordinate map denoted as CoordMapHR, and finally subtract CoordMapHR-Sampling to get coordinate offset denoted as Coordoffset, as shown in Eqs (9)-(10):

\[
\text{CoordMap}_{HR-Sampling} = \text{UpSampling(CoordMap}_{LR}) \tag{9}
\]

\[
\text{Coordoffset} = \text{CoordMap}_{HR} - \text{CoordMap}_{HR-Sampling} \tag{10}
\]

Then, the nearest neighbor interpolation is applied to Fepl,global for HR sampling epipolar geometry implicit code representation (i.e., FH,HR,global).

Finally, we concatenate the coordinate offset and HR sampling epipolar geometry implicit code as input to the multilayer perceptron f which consists of four linear layers with 256 output channels and one linear layer with 3 output channels to get HR (high resolution) EPI denoted as EPIHR, as shown in Eq (11):

\[
EPI_{HR} = f(Coord_{offset}, FH,HR,_{epi,global}) \tag{11}
\]

3.6. Divide-and-Conquer Learning Strategy

To reconstruct dense sampled light fields from sparse views, we need to train an implicit epipolar geometric function composed of MLPs. However, it is difficult to directly learn a 4D implicit epipolar geometric function because of limited resources and efficiency. Moreover, directly training a 4D implicit function may destroy the spatial and angular constraints of light field, because of the coupling relationship between the spatial and angular information of the light field. Therefore, we propose a divide-and-conquer learning strategy. Specifically, we divide the 4D implicit epipolar geometric function into two 2D implicit functions learning by fixing (u, s) and (v, t) respectively. When we fix (u, s), we can represent the light field as L(uvs, t), which corresponds to the horizontal EPI. Similarly, when we fix (v, t), the light field can be represented as L(u, vs, ti), which correspond to the vertical EPI. By taking the horizontal or vertical EPI tuples as input, we can reconstruct the light field in the v or u dimensions.

As shown in Figure 5, we reconstruct high angular resolution light field (7 × 7) from low angular resolution light field (3 × 3) through a three-stage strategy.

4. Experiments

We compare our method with existing methods on the 3 × 3 to 7 × 7 and 3 × 3 to 9 × 9 light field angular super-resolution tasks, including the comparisons of PSNR and SSIM metrics on synthetic and realistic datasets, as well as visual comparison of experimental results. Our network is trained using charbonnier loss [2] and is implemented in PyTorch on a PC with one NVidia TITAN X GPU.

4.1. 3 × 3 to 7 × 7 task

We use 100 scenes [13] as the training dataset. All scenes from 100scenes [13] are captured by Lytro camera [7] and decoded as a light field array image of 14 × 14. We take the middle 7 × 7 images as the light field data in our experiments. Since the generation of the entire light field includes horizontal EPI super-resolution stages and vertical EPI super-resolution stages, we recombine the sub-aperture images with a 7 × 7 array of each scene into a set of vertical and horizontal EPI tuples, taking the same size and shuffling to improve the generalization performance of the model when dealing with horizontal and vertical EPI.

We compare our method with 4 state of the art methods on 30scenes, occlusions, and reflective for the 3 × 3 to 7 × 7 tasks, including Kalantari [13], EPICNN [31],
Table 1. PSNR/SSIM achieved by different methods for $3 \times 3 \rightarrow 7 \times 7$ angular SR. The best results are shown in bold.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Task</th>
<th>30scenes</th>
<th>occlusions</th>
<th>reflective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kalantari [13]</td>
<td>3x3→7x7</td>
<td>38.41/0.9601</td>
<td>33.21/0.9012</td>
<td>36.89/0.9462</td>
</tr>
<tr>
<td>EPICNN [31]</td>
<td>3x3→7x7</td>
<td>41.04/0.9782</td>
<td>36.85/0.9382</td>
<td>41.24/0.9703</td>
</tr>
<tr>
<td>EP4DCNN [27]</td>
<td>3x3→7x7</td>
<td>43.82/0.9926</td>
<td>34.69/0.9231</td>
<td>39.93/0.9594</td>
</tr>
<tr>
<td>DistgASR [24]</td>
<td>3x3→7x7</td>
<td>44.18/0.9912</td>
<td>42.21/0.9949</td>
<td>41.75/0.9757</td>
</tr>
<tr>
<td>LFEIFASR(ours)</td>
<td>3x3→7x7</td>
<td><strong>44.46/0.9917</strong></td>
<td><strong>42.34/0.9902</strong></td>
<td><strong>41.77/0.9677</strong></td>
</tr>
</tbody>
</table>

Table 2. PSNR/SSIM achieved by different methods for $3 \times 3 \rightarrow 9 \times 9$ angular SR. The best results are shown in bold.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Task</th>
<th>UrbanLF-Syn</th>
<th>HCI old</th>
<th>HCI new</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kalantari [13]</td>
<td>3x3→9x9</td>
<td>36.78/0.9455</td>
<td>32.76/0.9028</td>
<td>26.59/0.8720</td>
</tr>
<tr>
<td>DistgASR [24]</td>
<td>3x3→9x9</td>
<td>43.43/0.9910</td>
<td>38.91/0.9594</td>
<td>31.93/0.9501</td>
</tr>
<tr>
<td>LFEIFASR(ours)</td>
<td>3x3→9x9</td>
<td><strong>43.55/0.9912</strong></td>
<td><strong>39.59/0.9444</strong></td>
<td><strong>32.43/0.9291</strong></td>
</tr>
</tbody>
</table>

4.2. $3 \times 3 \rightarrow 9 \times 9$ task

The $3 \times 3 \rightarrow 9 \times 9$ task is compared with existing methods on UrbanLF-Syn [22], HCI old [29], and HCI new [10], and the PSNR and SSIM obtained for each method in the dataset were calculated as shown in the Table 2.

LFEIFASR achieves the highest PSNR on all datasets. In terms of specific PSNR metrics, our model outperforms DistgASR by 0.28 dB on 30scenes dataset, by 0.13 dB on occlusions dataset, and by 0.02 dB on reflective dataset.

Figure 6 shows the visual comparison results on $3 \times 3$ to $7 \times 7$ task. Our model shows the best rendering quality and detail restoration effect in these scenes. For example, the figures on the scene IMG_1528 (Figure 6(a)), plants on the scene buddha2 (Figure 6(b)), realistic results in these scenes, such as the letters "HCI" on the scene dishes (Figure 7(c)), and the number "1851" on the scene dishes (Figure 7(c)).

4.3. Light field rendering with arbitrary angular resolution

Compared to other super-resolution with fixed angular resolution, our method can reconstruct light field images with higher angular resolution based on the input high-resolution coordinate. We take the $3 \times 3$ to $17 \times 17$, $3 \times 3$ to $25 \times 25$, $3 \times 3$ to $33 \times 33$, and $3 \times 3$ to $41 \times 41$ angular SR tasks for discussion. Since the adopted light field images on all datasets, and the highest SSIM value in 30 scenes.
As shown in Figure 8, LFEIFASR can still generate high-quality subaperture images when reconstructing higher angular resolution light field. LFEIFASR is able to maintain the epipolar geometry constraint in the light field using epipolar geometry convolution.

### 5. Ablation experiments

This section we discuss the influence of $m$ and epipolar geometric convolution EPIConv in LFEIFASR.

#### 5.1. The value of $m$ taken for the neighborhood EPIs of the light field

We test LFEIFASR with different values of $m$ for the light field neighborhood EPIs in 30 scenes. Different $m$ value will get different test results. As shown in Table 4, the highest PSNR and SSIM achieved at $m = 2$. (i.e. the light field EPI image is taken as an array of 5 EPI images).

Table 4. Effect of different values of $m$ (PSNR↑/SSIM↑)

<table>
<thead>
<tr>
<th>Values of $m$</th>
<th>Task</th>
<th>30 scenes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m=0$</td>
<td>3x3→7x7</td>
<td>43.89/0.9901</td>
</tr>
<tr>
<td>$m=1$</td>
<td>3x3→7x7</td>
<td>44.44/0.9917</td>
</tr>
<tr>
<td>$m=2$</td>
<td>3x3→7x7</td>
<td><strong>44.46/0.9917</strong></td>
</tr>
<tr>
<td>$m=3$</td>
<td>3x3→7x7</td>
<td>44.45/0.9917</td>
</tr>
<tr>
<td>$m=4$</td>
<td>3x3→7x7</td>
<td>44.39/0.9916</td>
</tr>
</tbody>
</table>
5.2. Comparison with residual dense network

The normal convolutional block is replaced by epipolar geometry convolution block to extract epipolar geometry information from the light field. We discuss the advantages of EPIBlock over RDB [37] in the light field angular super-resolution. 30scenes is used as test dataset, as shown in Table 5. It can be concluded that the LFEIFASR with epipolar geometric convolution block EPIBlock can obtain higher accuracy than the light field angular continuous domain plotting network with ordinary convolution block RDB, and reduce the number of parameters of the network model by about 40% at the same time.

### Table 5. Effect of normal convolution block and epipolar geometric convolution block in 30scenes (PSNR↑/SSIM↑).

<table>
<thead>
<tr>
<th>Type</th>
<th>Task</th>
<th>30scenes</th>
<th>Params</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDB</td>
<td>3x3→7x7</td>
<td>44.34/0.9911</td>
<td>22.4M</td>
</tr>
<tr>
<td>EPIBlock</td>
<td>3x3→7x7</td>
<td>44.46/0.9917</td>
<td>13.7M</td>
</tr>
</tbody>
</table>

6. Conclusions

In this paper, we propose a framework for learning continuous angular domain of light fields using light field epipolar geometry information. The framework learn the constrained properties of the epipolar geometry in light fields by a four-dimensional light field implicit epipolar geometric function. We decompose the learning process of the 4D implicit function into two sub-processes to solve the problem that learning the 4D implicit function is hard. We use a divide-and-conquer learning strategy to learn two two-dimensional implicit functions by fixing one angular coordinate and one spatial coordinate in the four-dimensional coordinates of the light field. In order to solve the problem of lack of spatial information after the fixed dimension, we take the multi-line EPI as the input of the network, and embed the spatial information into the angular information.

The experimental results demonstrate that we not only can synthesize high-quality LF image in low-magnification angular super-resolution tasks, but still achieve state-of-the-art performance in arbitrary and high-magnification light field reconstruction.

7. Acknowledgments

This work is supported in part by Joint Funds of the Zhejiang Provincial Natural Science Foundation of China (LTY22F020001), and the “Pioneer” and “Leading Goose” R&D Program of Zhejiang (2022C01005, 2022C01082).

### References


[31] Gaochang Wu, Yebin Liu, Lu Fang, Qionghai Dai, and Tianyou Chai. Light field reconstruction using convolutional network on EPI and extended applications. IEEE transactions on pattern analysis and machine intelligence, 41(7):1681–1694, 2018. 1, 2, 5, 6


