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].

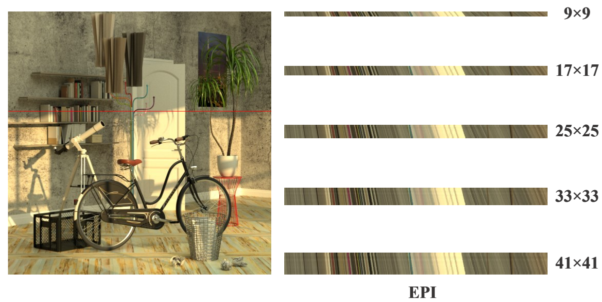


Figure 1. Implicit epipolar geometric function can preserve clear light field epipolar geometric structure in arbitrary angular resolution light field reconstruction.

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 7×7 or 9×9), 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,

*Bangcheng Zong and Lin Zhong contributed equally to this work and should be considered co-first authors

†Corresponding author: zhouwenhui@hdu.edu.cn

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 codes. To our best knowledge, our method is the first implicit function-based light field angular super-resolution.

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 SAIs, 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

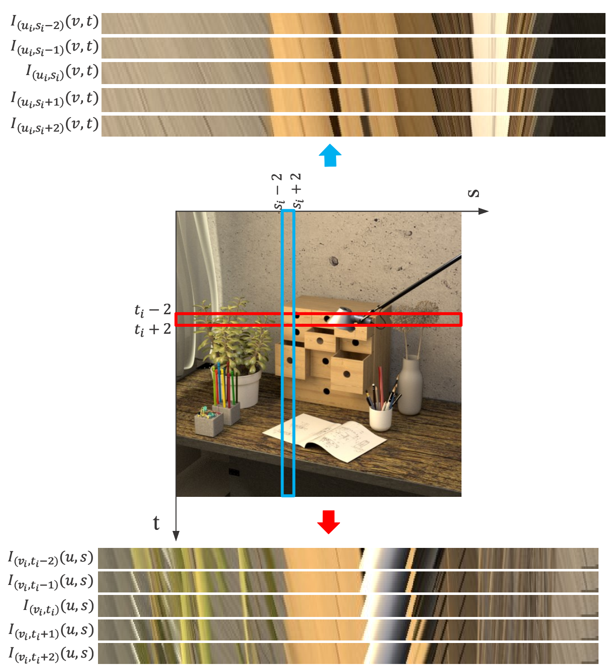


Figure 2. Examples of a horizontal neighborhood EPI array and a vertical neighborhood EPI array, here $m = 2$.

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 SAIs.

Specifically, as for a two-plane parameterized light field $I\left(u,v,s,t\right)$ with resolution of $U \times V \times W \times H$, a horizontal EPI $I_{(v_i,t_i)}\left(u,s\right)$ can be obtained by stacking the t_i-th row pixels of the v_i-th row SAIs, and its horizontal neighborhood EPI $I_{neib(v_i,t_i)}\left(u,s\right)$ is built by stitching $2m+1$ adjacent horizontal EPIs together along the u axis, including $I_{(v_i,t_{i-m})}\left(u,s\right)$, ..., $I_{(v_i,t_{i-1})}\left(u,s\right)$, $I_{(v_i,t_i)}\left(u,s\right)$, $I_{(v_i,t_{i+1})}\left(u,s\right)$, ..., $I_{(v_i,t_{i+m})}\left(u,s\right)$. The resolution of a horizontal neighborhood EPI is $(2m+1)U \times W$. Similarly, a vertical EPI is defined as $I_{(u_i,s_i)}\left(v,t\right)$, and its vertical neighborhood EPI $I_{neib(u_i,s_i)}\left(v,t\right)$ can be built by stitching $2m+1$ adjacent vertical EPIs together along the v axis, including $I_{(u_i,s_{i-m})}\left(v,t\right)$, ..., $I_{(u_i,s_{i-1})}\left(v,t\right)$, $I_{(u_i,s_i)}\left(v,t\right)$, $I_{(u_i,s_{i+1})}\left(v,t\right)$, ..., $I_{(u_i,s_{i+m})}\left(v,t\right)$. 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 neighborhood ones) 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 3×9 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.

$$\begin{aligned} F_{mixed} = &H_{Spatial\_Conv}(I_{epi\_neib}) \\ &+ H_{Angular\_Conv}(I_{epi\_basis}) \end{aligned} \tag{1}$$

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 1×7 convolution. Then the extracted feature is summed and fused by a 3×3 convlution

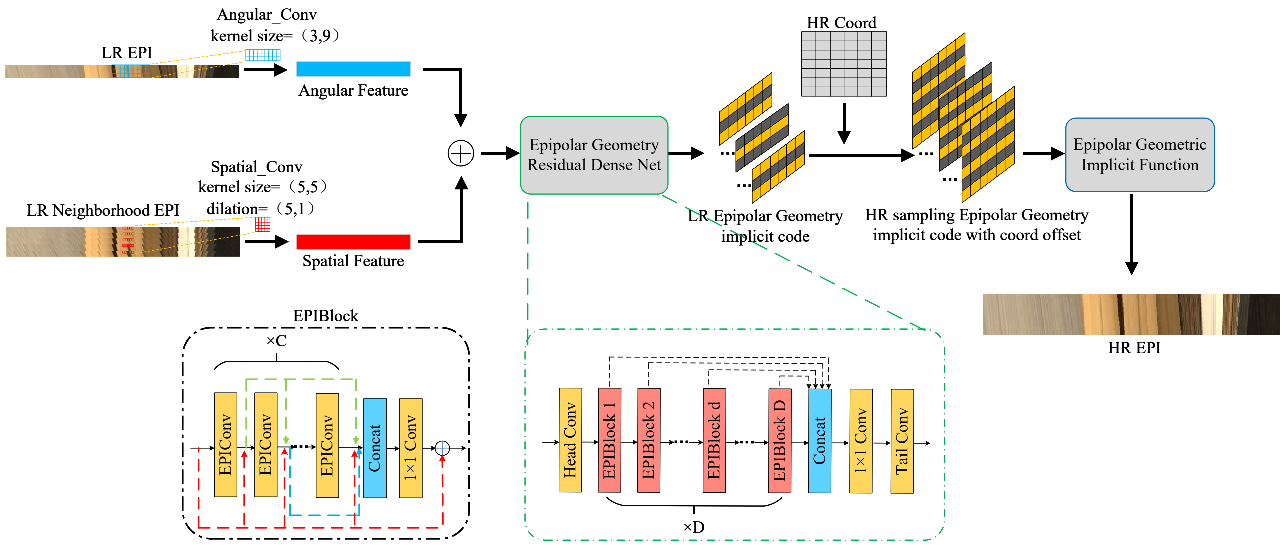


Figure 3. Overview of the proposed network architecture.

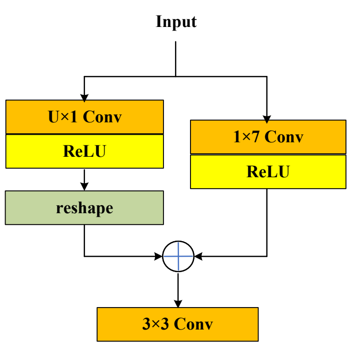


Figure 4. Structure of EPIConv.

to learn the slope information in EPIs, as shown in Eqs (2)-(5):

$$EPIConv(F) = Conv_{3\times3}(ReLU(Conv_{U\times1}(F)) + ReLU(Conv_{1\times7}(F))) \tag{2}$$

$$F_{d,c} = EPIConv(Concat(F_{d\text{-}1}, F_{d,1}, ..., F_{d,c\text{-}1})) \tag{3}$$

$$F_{d,epi} = H^{d}_{1\times1}(Concat(F_{d\text{-}1}, F_{d,1}, ..., F_{d,c}, ..., F_{d,C})) \tag{4}$$

$$F_d = F_{d\text{-}1} + F_{d,epi} \tag{5}$$

where $F_{d,c}$ denotes the epipolar geometric features obtained after c-layer EPIConv in the dth EPIBlock, $F_{d,epi}$ denotes the obtained features after all local convolutions, and F_d is the final output epipolar geometric features of the dth EPIBlock.

Base on EPIBlock and EPIConv, we use a 3×3 convolution denoted as $Head\_Conv$ to extract shallow features first. The spatial-angular fusion information F_{mixed} extracted by the spatial-angular information fusion module will be entered into $Head\_Conv$ to extract shallow features, as described in Eq (6):

$$F_0 = H_{Head\_Conv}(F_{mixed}) \tag{6}$$

Our epipolar geometry residuals dense network consists of D EPIBlocks, and each EPIBlock comprising of C EPIConvs, which forms a DenseNet structure to fully extract epipolar geometric features, that is each EPIConv accepts feature information from all previous EPIConvs. 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 EPIConvs are set to be 64, as shown in Eqs (7)-(8):

$$F_d = H_{EPIBlock,d}(F_{d\text{-}1}) = H_{EPIBlock,d}(H_{EPIBlock,d\text{-}1}(...(H_{EPIBlock,1}(F_0)))) \tag{7}$$

$$F_{epi\_global} = H_{TailConv}(F_{\text{-}1}, \quad F_0, F_1, ..., F_D) \tag{8}$$

where F_d denotes the epipolar geometry feature produced by the $(d-1)$th EPIBlock. The epipolar geometry global feature denoted as F_{epi_global} in the EPIs can be extracted by applying a 3×3 convolution (denoted as $TailConv$) to a tensor concatenating output from each EPIBlock, the F_{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., $CoordMap_{LR}$) with LR epipolar geometry implicit code (i.e., F_{epi_global}), and then the high angular resolution (HR) sampling coordinate map (i.e., $CoordMap_{HR-Sampling}$) is obtained from $CoordMap_{LR}$ through the nearest neighbor interpolation. The HR coordinate (HR Coord) is used to generate high angular resolution coordinate map denoted as $CoordMap_{HR}$, and finally subtract $CoordMap_{HR-Sampling}$ to get coordinate offset denoted as $Coord_{offset}$, as shown in Eqs (9)-(10):

$$\begin{aligned} CoordMap_{HR-Sampling} = \\ UpSampling(CoordMap_{LR}) \end{aligned} \tag{9}$$

$$\begin{aligned} Coord_{offset} = CoordMap_{HR} - \\ CoordMap_{HR-Sampling} \end{aligned} \tag{10}$$

Then, the nearest neighbor interpolation is applied to F_{epi_global} for HR sampling epipolar geometry implicit code representation (i.e., $F_{HR\_epi\_global}$).

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 EPI_{HR}, as shown in Eq (11):

$$EPI_{HR} = f(Coord_{offset}, F_{HR\_epi\_global}) \tag{11}$$

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 $L(u_i, v, s_i, t)$, which corresponds to the horizontal EPI. Similarly, when we fix (v, t), the light field can be represented as $L(u, v_i, s, t_i)$, 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.

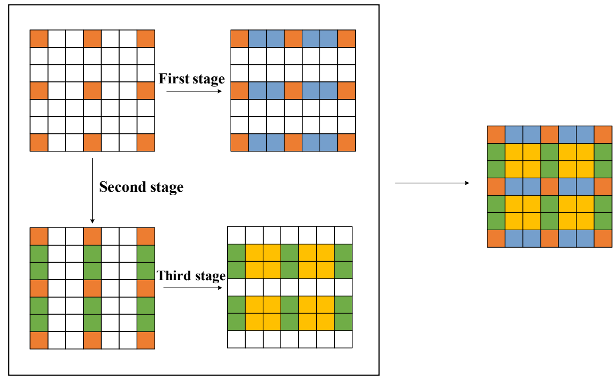


Figure 5. Synthesis of high resolution light field. We take sparse viewpoints marked in orange as input first, use the horizontal EPI to synthesize the viewpoints marked in blue, and use the vertical EPI to synthesize the viewpoints marked in green. Secondly, we use the green viewpoint as input, and use the horizontal EPI to synthesize the viewpoints marked in yellow.

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],

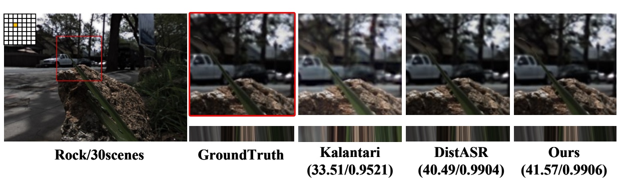
Table 1. PSNR/SSIM achieved by different methods for $3 \times 3 \rightarrow 7 \times 7$ angular SR. The best results are shown in bold.

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

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.

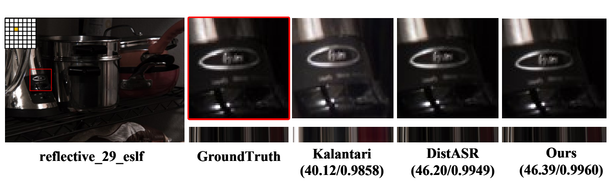
Methods	Task	UrbanLF-Syn	HCI old	HCI new
Kalantari [13]	3x3→9x9	36.78/0.9455	32.76/0.9028	26.59/0.8720
DistgASR [24]	3x3→9x9	43.43/0.9910	38.91/**0.9659**	31.93/**0.9501**
LFEIFASR(ours)	3x3→9x9	**43.55**/**0.9912**	**39.59**/0.9444	**32.43**/0.9291

(a) IMG_1528 in 30scenes



(b) Rock in 30scenes

(c) occlusions_24 in occlusions



(d) reflective_29 in reflective

Figure 6. Visualization comparison of LFEIFASR and other published methods in 3×3 to 7×7 task

EP4DCNN [27] and DistgASR [24]. Table 1 lists the comparison results. Our model achieves the highest PSNR values on all datasets, and the highest SSIM value in 30 scenes. 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 3×3 to 7×7 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 $Rock$ (Figure 6(b)).

4.2. 3×3 to 9×9 task

The 3×3 to 9×9 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.

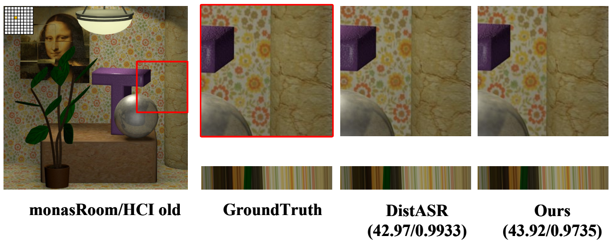
LFEIFASR achieves the highest PSNR on all datasets. In terms of specific PSNR metrics, there is 0.12 dB higher than the latest method DistgASR [24] on UrbanLF-Syn [22], 0.68 dB higher than DistgASR on HCI old [29], and 0.50 dB higher than DistgASR on HCI new [10].

Figure 7 shows the reconstruction results of LFEIFASR in some scenes from 3×3 to 9×9 tasks and the comparison with DistgASR. LFEIFASR reconstruct more photo-realistic results in these scenes, such as the letters "HCI" on the scene buddha2 (Figure 7(a)) and the number "1851" on the scene dishes (Figure 7(c)).

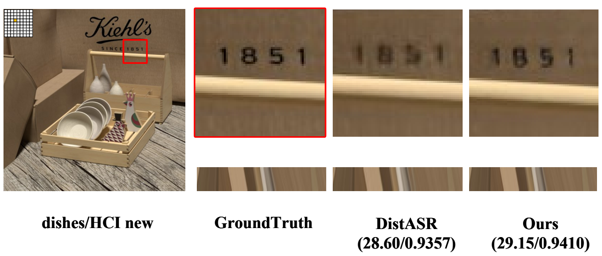
4.3. Light field rendering with arbitrary angular resolution

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 3×3 to 17×17, 3×3 to 25×25, 3×3 to 33×33, and 3×3 to 41×41 angular SR tasks for discussion. Since the adopted light field

(a) buddha2 in HCI old



(b) monasRoom in HCI old



(c) dishes in HCI new

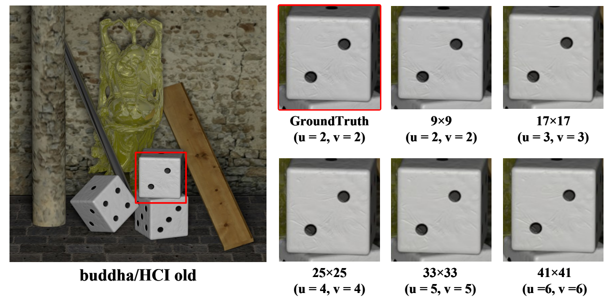
Figure 7. Visualization comparison of LFEIFASR and other published methods in 3x3 to 9x9 task

Table 3. PSNR/SSIM of our method in higher angular SR tasks.

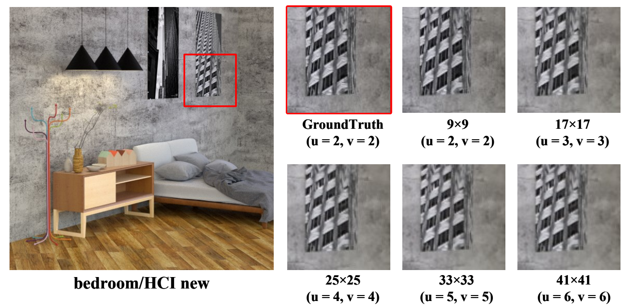
Tasks	Bedroom	Bicyle	Buddha
3x3→9x9	37.52/0.9664	35.10/0.9694	46.27/0.9945
3x3→17x17	36.12/0.9602	34.16/0.9631	43.37/0.9918
3x3→25x25	35.32/0.9533	33.59/0.9570	42.29/0.9892
3x3→33x33	34.89/0.9487	33.28/0.9531	41.77/0.9877
3x3→41x41	34.63/0.9456	33.10/0.9505	41.46/0.9867

dataset only has an angular resolution of 9×9 and there is no ground-truth with 17×17 and higher angular resolutions, we sample a regular 9×9 subaperture image array in a higher angular resolution light field to calculate the evaluation metrics when testing the synthesis result. We use the scene $bedroom$ and $bicyle$ on HCI new and $buddha$ on HCI old for evaluation.

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.



(a) Performance of buddha scene in HCI old dataset with arbitrary angular resolution light field drawin



(b) Performance of bedroom scenes in the HCI new dataset at any angular resolution light field drawing

Figure 8. Experimental results of LFEIFASR with multiple target angular resolutions under arbitrary angular resolution light field mapping

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 30scenes. 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↑)

Values of m	Task	30scenes
m=0	3x3→7x7	43.89/0.9901
m=1	3x3→7x7	44.44/0.9917
m=2	3x3→7x7	**44.46/0.9917**
m=3	3x3→7x7	44.45/0.9917
m=4	3x3→7x7	44.39/0.9916

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↑).

Type	Task	30scenes	Params
RDB	3×3→7×7	44.34/0.9911	22.4M
EPIBlock	3×3→7×7	**44.46/0.9917**	**13.7M**

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

[1] Vivek Boominathan, Kaushik Mitra, and Ashok Veeraraghavan. Improving resolution and depth-of-field of light field cameras using a hybrid imaging system. In *2014 IEEE International Conference on Computational Photography (ICCP)*, pages 1–10. IEEE, 2014. 1

[2] Pierre Charbonnier, Laure Blanc-Feraud, Gilles Aubert, and Michel Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proceedings of 1st international conference on image processing*, volume 2, pages 168–172. IEEE, 1994. 5

[3] Yinbo Chen, Sifei Liu, and Xiaolong Wang. Learning continuous image representation with local implicit image function. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8628–8638, 2021. 1, 2, 5

[4] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5939–5948, 2019. 2

[5] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, pages 184–199. Springer, 2014. 5

[6] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 391–407. Springer, 2016. 5

[7] Todor Georgiev, Zhan Yu, Andrew Lumsdaine, and Sergio Goma. Lytro camera technology: theory, algorithms, performance analysis. In *Multimedia content and mobile devices*, volume 8667, pages 458–467. SPIE, 2013. 5

[8] M Shahzeb Khan Gul and Bahadir K Gunturk. Spatial and angular resolution enhancement of light fields using convolutional neural networks. *IEEE Transactions on Image Processing*, 27(5):2146–2159, 2018. 1, 2

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 3

[10] Katrin Honauer, Ole Johannsen, Daniel Kondermann, and Bastian Goldluecke. A dataset and evaluation methodology for depth estimation on 4d light fields. In *Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part III 13*, pages 19–34. Springer, 2017. 6

[11] Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of PSNR in image/video quality assessment. *Electronics letters*, 44(13):800–801, 2008. 1

[12] Ivo Ihrke, John Restrepo, and Lois Mignard-Debise. Principles of light field imaging: Briefly revisiting 25 years of research. *IEEE Signal Processing Magazine*, 33(5):59–69, 2016. 1

[13] Nima Khademi Kalantari, Ting-Chun Wang, and Ravi Ramamoorthi. Learning-based view synthesis for light field cameras. *ACM Transactions on Graphics (TOG)*, 35(6):1–10, 2016. 1, 5, 6

[14] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1637–1645, 2016. 5

[15] Keunsoo Ko, Yeong Jun Koh, Soonkeun Chang, and Chang-Su Kim. Light field super-resolution via adaptive feature remixing. *IEEE Transactions on Image Processing*, 30:4114–4128, 2021. 2

[16] Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42, 1996. 1

[17] Jiaxin Li, Zijian Feng, Qi She, Henghui Ding, Changhu Wang, and Gim Hee Lee. MINE: Towards continuous depth mpi with nerf for novel view synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12578–12588, 2021. 2

[18] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 1, 2

[19] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan. *Light field photography with a hand-held plenoptic camera*. PhD thesis, Stanford University, 2005. 1

[20] Sida Peng, Yuanqing Zhang, Yinghao Xu, Qianqian Wang, Qing Shuai, Hujun Bao, and Xiaowei Zhou. Neural body: Implicit neural representations with structured latent codes for novel view synthesis of dynamic humans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9054–9063, 2021. 2

[21] Mehdi SM Sajjadi, Bernhard Scholkopf, and Michael Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. In *Proceedings of the IEEE international conference on computer vision*, pages 4491–4500, 2017. 5

[22] Hao Sheng, Ruixuan Cong, Da Yang, Rongshan Chen, Sizhe Wang, and Zhenglong Cui. Urbanlf: a comprehensive light field dataset for semantic segmentation of urban scenes. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(11):7880–7893, 2022. 6

[23] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems*, 33:7462–7473, 2020. 2

[24] Yingqian Wang, Longguang Wang, Gaochang Wu, Jungang Yang, Wei An, Jingyi Yu, and Yulan Guo. Disentangling light fields for super-resolution and disparity estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):425–443, 2022. 2, 6

[25] Yunlong Wang, Fei Liu, Zilei Wang, Guangqi Hou, Zhenan Sun, and Tieniu Tan. End-to-end view synthesis for light field imaging with pseudo 4DCNN. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 333–348, 2018. 2

[26] Yunlong Wang, Fei Liu, Kunbo Zhang, Guangqi Hou, Zhenan Sun, and Tieniu Tan. LFNet: A novel bidirectional recurrent convolutional neural network for light-field image super-resolution. *IEEE Transactions on Image Processing*, 27(9):4274–4286, 2018. 2

[27] Yunlong Wang, Fei Liu, Kunbo Zhang, Zilei Wang, Zhenan Sun, and Tieniu Tan. High-fidelity view synthesis for light field imaging with extended pseudo 4DCNN. *IEEE Transactions on Computational Imaging*, 6:830–842, 2020. 2, 6

[28] Sven Wanner and Bastian Goldluecke. Variational light field analysis for disparity estimation and super-resolution. *IEEE transactions on pattern analysis and machine intelligence*, 36(3):606–619, 2013. 1

[29] Sven Wanner, Stephan Meister, and Bastian Goldluecke. Datasets and benchmarks for densely sampled 4d light fields. In *VMV*, volume 13, pages 225–226, 2013. 6

[30] Gaochang Wu, Yebin Liu, Qionghai Dai, and Tianyou Chai. Learning sheared EPI structure for light field reconstruction. *IEEE Transactions on Image Processing*, 28(7):3261–3273, 2019. 2

[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

[32] Gaochang Wu, Belen Masia, Adrian Jarabo, Yuchen Zhang, Liangyong Wang, Qionghai Dai, Tianyou Chai, and Yebin Liu. Light field image processing: An overview. *IEEE Journal of Selected Topics in Signal Processing*, 11(7):926–954, 2017. 1, 2

[33] Henry Wing Fung Yeung, Junhui Hou, Xiaoming Chen, Jie Chen, Zhibo Chen, and Yuk Ying Chung. Light field spatial super-resolution using deep efficient spatial-angular separable convolution. *IEEE Transactions on Image Processing*, 28(5):2319–2330, 2018. 2

[34] Wang Yifan, Shihao Wu, Cengiz Oztireli, and Olga Sorkine-Hornung. Iso-Points: Optimizing neural implicit surfaces with hybrid representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 374–383, 2021. 2

[35] Youngjin Yoon, Hae-Gon Jeon, Donggeun Yoo, Joon-Young Lee, and In So Kweon. Learning a deep convolutional network for light-field image super-resolution. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 24–32, 2015. 1, 2

[36] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelNerf: Neural radiance fields from one or few images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4578–4587, 2021. 1, 2

[37] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481, 2018. 3, 8