FAST: FLOW-ASSISTED SHEARLET TRANSFORM FOR DENSELY-SAMPLED LIGHT FIELD RECONSTRUCTION

Yuan Gao, Reinhard Koch

Kiel University, Germany

{yga, rk}@informatik.uni-kiel.de

Robert Bregovic, Atanas Gotchev

Tampere University, Finland

{robert.bregovic, atanas.gotchev}@tuni.fi

ABSTRACT

Shearlet Transform (ST) is one of the most effective methods for Densely-Sampled Light Field (DSLF) reconstruction from a Sparsely-Sampled Light Field (SSLF). However, ST requires a precise disparity estimation of the SSLF. To this end, in this paper a state-of-the-art optical flow method, *i.e*. PWC-Net, is employed to estimate bidirectional disparity maps between neighboring views in the SSLF. Moreover, to take full advantage of optical flow and ST for DSLF reconstruction, a novel learning-based method, referred to as Flow-Assisted Shearlet Transform (FAST), is proposed in this paper. Specifically, FAST consists of two deep convolutional neural networks, *i.e*. disparity refinement network and view synthesis network, which fully leverage the disparity information to synthesize novel views via warping and blending and to improve the novel view synthesis performance of ST. Experimental results demonstrate the superiority of the proposed FAST method over the other state-of-the-art DSLF reconstruction methods on nine challenging real-world SSLF sub-datasets with large disparity ranges (up to 26 pixels).

Index Terms— Densely-Sampled Light Field Reconstruction, Parallax View Generation, Novel View Synthesis, Shearlet Transform, Flow-Assisted Shearlet Transform

1. INTRODUCTION

Densely-Sampled Light Field (DSLF) is a discrete representation of the 4D approximation of the plenoptic function parameterized by two parallel planes (camera plane and image plane) [1], where multi-perspective camera views are arranged in such a way that the disparities between adjacent views are less than one pixel [2]. How to reconstruct a DSLF from a Sparsely-Sampled Light Field (SSLF) is depicted in Fig. 1. The solid-line orange cameras, *i.e*. $\mathcal{C}_j$, are uniformly distributed along the horizontal axis with the same camera orientation and focal length. The images captured by them for a static scene compose a horizontal-parallax SSLF. The DSLF reconstruction on this 3D SSLF can be considered as novel view synthesis between any two neighboring parallax views in this SSLF, of which the results are represented by the images from the dash-line blue cameras in Fig. 1. Let the interpolation rate of this novel view synthesis process be denoted by δ, it is apparent that the virtual camera $\mathcal{C}_{j+t}$ meets the condition that $t \in \{\frac{1}{\delta}, \frac{2}{\delta}, ..., \frac{\delta-1}{\delta}\}$. Besides, in order to reconstruct the target unknown DSLF correctly, δ should be greater than the disparity range of the input SSLF (see Sect. 4.1).

Shearlet Transform (ST) [3, 4] is especially effective in reconstructing a DSLF from a SSLF with a large disparity range (> 16 pixels). The disparity information of the SSLF is required to be obtained in advance for 1) constructing a decent shearlet system; 2) pre-shearing the SSLF in order to eliminate the minimum disparity of it. To tackle the disparity-estimation problem, a state-of-the-art optical flow algorithm, *i.e*. PWC-Net [5], is exploited for estimating the bidirectional disparity maps between adjacent views in the SSLF.

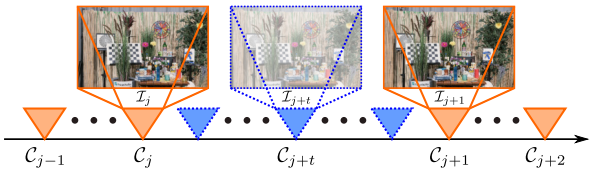


Fig. 1. DSLF reconstruction on a SSLF.

In addition to assisting the DSLF reconstruction of using ST, the estimated bidirectional disparity maps can also be used to perform DSLF reconstruction via novel view synthesis using image warping and blending techniques. However, due to the occlusion and errors in the estimated disparity maps, this disparity-based solution to DSLF reconstruction may not produce visually pleasing results. Therefore, to improve the performance of both ST-based and disparity-based DSLF reconstruction methods, a novel learning-based approach, referred to as Flow-Assisted Shearlet Transform (FAST), is proposed in this paper. The FAST method makes full use of the bidirectional disparity maps predicted by PWC-Net and the DSLF recovered by ST to better reconstruct the target DSLF via two deep convolutional neural networks, *i.e*. disparity refinement network and view synthesis network. Additionally, the proposed FAST is fully convolutional and end-to-end trainable. Experimental results on nine evaluation DSLF sub-datasets demonstrate the effectiveness of FAST for reconstructing DSLFs from SSLFs with large disparity ranges.

2. RELATED WORK

Learning-based novel view synthesis. Niklaus *et al*. interpolate a novel frame between two consecutive video frames using a deep fully Convolutional Neural Network (CNN), which estimates a spatially-adaptive convolution kernel that captures both motion and re-sampling coefficients for each pixel [6]. Due to the large memory demand of outputting all the 2D kernels for all the image pixels, a spatially-adaptive Separable Convolution (SepConv) method is proposed by replacing a 2D kernel with two 1D kernels, thus effectively reducing the memory requirement [7]. Jiang *et al*. propose a CNN-based variable-length multi-frame interpolation method, consisting of a flow computation CNN and a flow interpolation CNN, for generating as many intermediate frames as needed between two consecutive video frames [8].

Light field angular super-resolution. Since a horizontal-parallax SSLF can be converted into a video captured by a virtual camera moving horizontally, Gao and Koch extend SepConv with enhanced kernels and propose Parallax-Interpolation Adaptive Separable Convolution (PIASC) for DSLF reconstruction [9]. In addition, a horizontal-parallax SSLF can also be represented by several Epipolar-Plane Images (EPIs), thus the DSLF reconstruction problem is equal to how to reconstruct densely-sampled EPIs from sparsely-sampled EPIs, which is effectively solved by ST in the shearlet transform domain via EPI sparse representation [10, 3, 4].

3. METHODOLOGY

3.1. Shearlet Transform (ST)

The ST approach is originally proposed in [3] and extended in [4] for addressing the DSLF reconstruction problem with varying disparities. The core idea of ST is to design an elaborately-tailored universal shearlet system [3, 11] and to perform regularization in the shearlet transform domain for EPIs via an iterative α-adaptive algorithm [3] or a double overrelaxation (DORE) algorithm [4]. The construction of the specifically-designed shearlet system relies on the precise disparity estimation of the input SSLF. In particular, the minimum disparity d_{min} and maximum disparity d_{max} of this SSLF are required to be estimated before applying the ST method. The corresponding disparity range of the input SSLF is determined by them, *i.e.* $d_{range} = d_{max} - d_{min}$. Based on the value of the estimated d_{min}, a pre-shearing process using cubic interpolation is then performed on the input SSLF, so that the new minimum disparity $d'_{min} = 0$, the new maximum disparity $d'_{max} = d_{range}$ and the sheared SSLF is capable of being correctly processed by a shearlet system with ξ scales, where $\xi = \lceil \log_2 d_{range} \rceil$. Finally, a post-processing shearing operation is applied to the reconstructed DSLF in order to compensate for the disparity shift that is caused by the pre-shearing step.

3.2. Image warping and blending using optical flow

The bidirectional optical flows between two video frames are effective in interpolating novels views between them via warping and blending [8]. Since a horizontal-parallax SSLF can be treated as a video captured by a virtual camera moving along the horizontal axis, optical flow is used here to solve the DSLF reconstruction problem. Assume $j = 0$ as illustrated in Fig. 1, the novel view $\mathcal{I}_t$ of $\mathcal{C}_t$ between $\mathcal{C}_0$ and $\mathcal{C}_1$ is synthesized by

$$\begin{aligned} \mathcal{I}_t &= \lambda \circ g(\mathcal{I}_0, \mathcal{F}_{t\to 0}) + (1-\lambda) \circ g(\mathcal{I}_1, \mathcal{F}_{t\to 1}), \\ \lambda &= \frac{(1-t)\mathcal{V}_{t\leftarrow 0}}{(1-t)\mathcal{V}_{t\leftarrow 0} + t(1-\mathcal{V}_{t\leftarrow 0})}, \end{aligned} \tag{1}$$

where $\mathcal{F}_{t\to 0}$ and $\mathcal{F}_{t\to 1}$ denote the optical flows from $\mathcal{I}_t$ to $\mathcal{I}_0$ and $\mathcal{I}_t$ to $\mathcal{I}_1$, $g(\cdot,\cdot)$ is an inverse warping function using bilinear interpolation, '$\circ$' denotes the element-wise (Hadamard) product and $\mathcal{V}_{t\leftarrow 0}$ represents the soft visibility map from $\mathcal{C}_0$ to $\mathcal{C}_t$. However, it is difficult to compute the inverse optical flows, *i.e.* $\mathcal{F}_{t\to 0}$ and $\mathcal{F}_{t\to 1}$ in (1), because the target novel view $\mathcal{I}_t$ is unknown. Since the bidirectional optical flows, *i.e.* $\mathcal{F}_{0\to 1}$ and $\mathcal{F}_{1\to 0}$, are much easier to be estimated, the inverse optical flows are typically approximated from them via

$$\begin{aligned} \tilde{\mathcal{F}}_{t\to 0} &= -(1-t)t\mathcal{F}_{0\to 1} + t^2\mathcal{F}_{1\to 0}, \\ \tilde{\mathcal{F}}_{t\to 1} &= (1-t)^2\mathcal{F}_{0\to 1} - t(1-t)\mathcal{F}_{1\to 0}. \end{aligned} \tag{2}$$

3.3. Flow-Assisted Shearlet Transform (FAST)

Inspired by the success of Super-SloMo [8] in video frame interpolation, a novel learning-based method, referred to as Flow-Assisted Shearlet Transform (FAST), is proposed to reconstruct DSLFs from SSLFs. The FAST method adopts a state-of-the-art optical flow approach, *i.e.* PWC-Net [5], to estimate the bidirectional optical flows between neighboring views in a SSLF. Besides, FAST also leverages a state-of-the-art DSLF reconstruction method, *i.e.* ST [4], to guide novel view synthesis. The architecture of FAST is illustrated in Fig. 2. As can be seen from this figure, FAST is composed of two deep convolutional neural networks based on the U-Net architecture [12], which are Disparity Refinement Network (DRN) and View Synthesis Network (VSN). Regarding the architecture of DRN, it has six hierarchies in the encoder part and five hierarchies in the decoder part with the same architecture as the flow interpolation CNN in Super-SloMo. Since the horizontal-parallax SSLF shown in Fig. 1 does not involve any object motion along the vertical axis, only the horizontal components of the bidirectional optical flows estimated by PWC-Net contain useful information, which are denoted by $\tilde{\mathcal{F}}^u_{0\to 1}$ and $\tilde{\mathcal{F}}^u_{1\to 0}$ and called bidirectional disparity maps. The inverse disparity maps, *i.e.* $\tilde{\mathcal{F}}^u_{t\to 0}$ and $\tilde{\mathcal{F}}^u_{t\to 1}$, can then be estimated by using (2). The DRN of FAST takes $\tilde{\mathcal{F}}^u_{0\to 1}$, $\tilde{\mathcal{F}}^u_{1\to 0}$, $\tilde{\mathcal{F}}^u_{t\to 0}$, $\tilde{\mathcal{F}}^u_{t\to 1}$, $g(\mathcal{I}_0, \tilde{\mathcal{F}}^u_{t\to 0})$, $g(\mathcal{I}_1, \tilde{\mathcal{F}}^u_{t\to 1})$, $\mathcal{I}_0$, $\mathcal{I}_1$ and $\mathcal{I}_t^{\text{ST}}$ as the input (19 channels in total) and outputs $\mathcal{F}^u_{t\to 0}$, $\mathcal{F}^u_{t\to 1}$ and $\mathcal{V}_{t\leftarrow 0}$, which are used to interpolate an intermediate view $\tilde{\mathcal{I}}_t$ via (1). With regard to the architecture of VSN, it is a "shallow" version of DRN with four hierarchies in the encoder part and three hierarchies in the decoder part. The interpolated novel view $\tilde{\mathcal{I}}_t$ and the corresponding view reconstructed by ST, *i.e.* $\mathcal{I}_t^{\text{ST}}$, are fed to the VSN of FAST to generate the final target view $\mathcal{I}_t$.

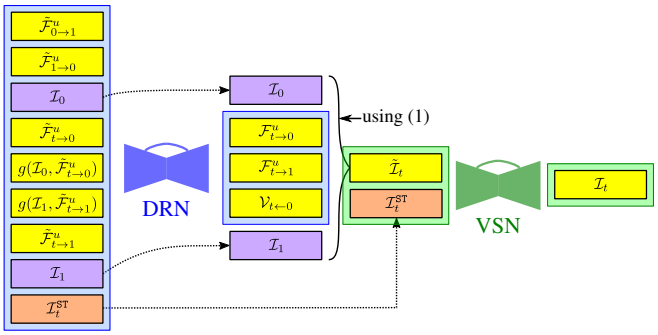


Fig. 2. An overview of the architecture of FAST. Note that $\mathcal{I}_t^{\text{ST}}$ is the view reconstructed by ST and $t \in \left\{\frac{1}{\delta}, \frac{2}{\delta}, ..., \frac{\delta-1}{\delta}\right\}$ (see Sect. 1).

Loss functions. The loss function of FAST is composed of VSN reconstruction loss, DRN reconstruction loss and warping loss, all of which are based on ℓ_1 norm:

$$\mathcal{L}^{\text{FAST}} = \omega_1\mathcal{L}^{\text{VSN}} + \omega_2\mathcal{L}^{\text{DRN}} + \omega_3\mathcal{L}^{\text{W}}, \tag{3}$$

where

$$\begin{aligned} \mathcal{L}^{\text{VSN}} &= \left\|\mathcal{I}_t - \mathcal{I}_t^{\text{GT}}\right\|_1, \\ \mathcal{L}^{\text{DRN}} &= \left\|\tilde{\mathcal{I}}_t - \mathcal{I}_t^{\text{GT}}\right\|_1, \\ \mathcal{L}^{\text{W}} &= \left\|g(\mathcal{I}_0, \mathcal{F}^u_{t\to 0}) - \mathcal{I}_t^{\text{GT}}\right\|_1 + \left\|g(\mathcal{I}_1, \mathcal{F}^u_{t\to 1}) - \mathcal{I}_t^{\text{GT}}\right\|_1, \end{aligned} \tag{4}$$

$\omega_1 = 9$, $\omega_2 = 2$ and $\omega_3 = 1$. Note that these weights are set empirically with a consideration that VSN reconstruction loss is more important than the other two losses.

4. EXPERIMENTS

4.1. Experimental settings

Training dataset. Two light field datasets are used for the training process. One is the Standford light field dataset captured by the Lego gantry with an angular resolution of 17×17. The other is the 4D light field benchmark with an angular resolution of 9×9 created with Blender [13]. The Standford Lego-gantry light field dataset is composed of 13 4D light fields. In each of these light fields, the center region with a size of 512×512 pixels *w.r.t.* each view is cut to make up 17 horizontal-parallax sub-datasets, each of which has 17 parallax images. Note that not all of these 13 4D light fields can be used for the training process, which is because 1) the PWC-Net algorithm fails in disparity estimation for the light field scenes containing reflective and transparent objects; 2) for some scenes, the estimated disparity range is beyond 64 pixels, which will be extremely expensive *w.r.t.* computation time if using ST method; 3) only static scenes are considered here. Therefore, eight 4D light fields are picked out of the Standford light field dataset. For the 4D light field benchmark, it is originally designed for depth estimation from 4D light fields. The

(a) $\mathcal{D}_1$: Books and charts (b) $\mathcal{D}_2$: Lego city (c) $\mathcal{D}_3$: Lightfield production (d) $\mathcal{D}_4$: Plants (e) $\mathcal{D}_5$: Table in the garden

(f) $\mathcal{D}_6$: Table top I (g) $\mathcal{D}_7$: Table top II (h) $\mathcal{D}_8$: Table top III (i) $\mathcal{D}_9$: Workshop (j) Cutting and scaling

Fig. 3. Middle views of evaluation DSLF sub-datasets $\mathcal{D}_\mu$ ($1 \leqslant \mu \leqslant 9$) and (j) illustrates the image cutting and scaling strategy.

benchmark contains four stratified, four test, and four training synthetic light field scenes. Besides, there are 16 additional community-supported photorealistic light field scenes captured in the same way as the benchmark. All of these 28 4D light fields have the same spatial resolution of 512×512 pixels and ground-truth d_{min} and d_{max}, which can facilitate the ST-based DSLF reconstruction. Each 4D light field is then converted into three horizontal-parallax sub-datasets, corresponding to 1st, 5th and 9th rows of views of it.

Training data augmentation. As introduced above, the training dataset consists of 220 horizontal-parallax sub-datasets, of which 136 have the resolution of $17 \times 512 \times 512 \times 3$ and the remaining ones have the resolution of $9 \times 512 \times 512 \times 3$. For each horizontal-parallax sub-dataset, nine continuous parallax images are randomly selected and cropped together into nine image patches with a size of 352×352 pixels during each iteration of the training process. The first and ninth cropped image patches are represented by $\mathcal{I}_0$ and $\mathcal{I}_1$, and used as input views for PWC-Net, ST and FAST. The remaining image patches are denoted by $\mathcal{I}_t$, $t \in \left\{\frac{1}{8}, \frac{2}{8}, ..., \frac{7}{8}\right\}$, which are used as the target views to be reconstructed for the neural network training of FAST.

Evaluation dataset. High Density Camera Array (HDCA) dataset is a real-world high-resolution 4D light field dataset captured by a DSLR camera mounted on a precisely-controlled gantry [14]. This dataset is composed of nine different 4D light fields with the same spatial resolution of 3976×2652 pixels. Eight of these light fields have the angular resolution of 101×21 and the remaining one has 99×21 views. Nevertheless, directly using the HDCA dataset is inappropriate for the performance evaluation of different DSLF reconstruction methods, because 1) all 4D light fields in this dataset are not densely-sampled; 2) black borders caused by calibration are not cut out as shown in Fig. 3 (j). Accordingly, a novel cutting and scaling strategy is proposed for transforming all the nine light fields of the HDCA dataset into DSLFs as illustrated in Fig. 3 (j). In particular, a $16:10$ image at the bottom center with occupying 90% of the width of the original view is cut and downscaled to a new resolution of 512×320 pixels by bicubic interpolation. Afterwards, only the top 97 horizontal-parallax views of each light field are used to construct a corresponding 3D DSLF, of which the resolution is $97 \times 512 \times 320 \times 3$. The generated ground-truth evaluation DSLF sub-datasets are denoted by $\mathcal{D}_\mu = \{\mathcal{I}^d_{i,\mu} | 1 \leqslant i \leqslant n\}$, where $\mathcal{I}^d_{i,\mu} \in \mathbb{R}^{M \times N \times 3}$, $1 \leqslant \mu \leqslant 9$, $n = 97$, $M = 512$ and $N = 320$. The middle views of them, *i.e.* $\mathcal{I}^d_{49,\mu}$, are shown in Fig. 3 (a)-(i). After setting the interpolation rate $\delta = 32$, the corresponding evaluation SSLF sub-datasets are represented by $\mathcal{S}_\mu = \{\mathcal{I}^s_{j,\mu} | 1 \leqslant j \leqslant m\}$, where $m = \frac{n-1}{\delta} + 1 (= 4)$. The estimated maximum and minimum disparities and disparity range of each $\mathcal{S}_\mu$ by PWC-Net are illustrated in Fig. 4. It can be found that the disparity-range values

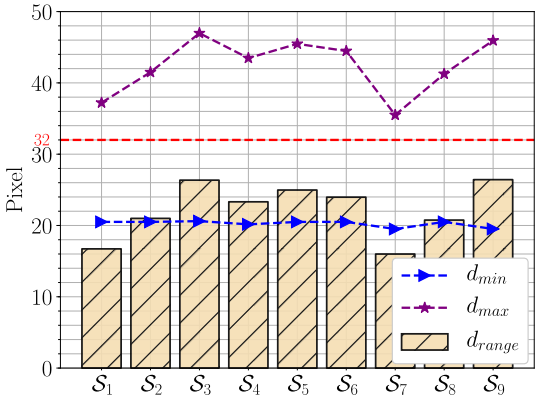


Fig. 4. Disparity estimations of $\mathcal{S}_\mu$ ($1 \leqslant \mu \leqslant 9$) using PWC-Net.

of the nine evaluation SSLF sub-datasets vary from 16 to 26 pixels; consequently, $\frac{d_{range}}{\delta} < 1$ pixel, which suggests that all the ground-truth DSLF sub-datasets $\mathcal{D}_\mu$ ($1 \leqslant \mu \leqslant 9$) are densely sampled.

Evaluation criteria. The per-view PSNR for a ground-truth DSLF sub-dataset $\mathcal{D}_\mu$ and the corresponding DSLF sub-dataset $\tilde{\mathcal{D}}_\mu$ reconstructed from $\mathcal{S}_\mu$ is described as below:

$$\begin{aligned} \mathrm{MSE}_{i,\mu} &= \frac{1}{3 \times M \times N} \sum_{x=1}^{M} \sum_{y=1}^{N} \left\| \tilde{\mathcal{I}}^d_{i,\mu}(x,y) - \mathcal{I}^d_{i,\mu}(x,y) \right\|_2^2, \\ \mathrm{PSNR}_{i,\mu} &= 10 \log_{10} \left(\frac{255^2}{\mathrm{MSE}_{i,\mu}} \right). \end{aligned} \tag{5}$$

The minimum per-view PSNR for each $\mathcal{D}_\mu$ is used as the final evaluation criteria as same as [9].

Implementation details. The proposed FAST method is implemented by using PyTorch, where the training mini-batch size is 6, the Adam optimizer is employed for training $1,500$ epochs and the learning rate of it is fixed to be 0.0001. The whole training process takes around 36 hours on an Nvidia Titan Xp GPU. Besides, the parameters of ST using the DORE algorithm are set as same as [4].

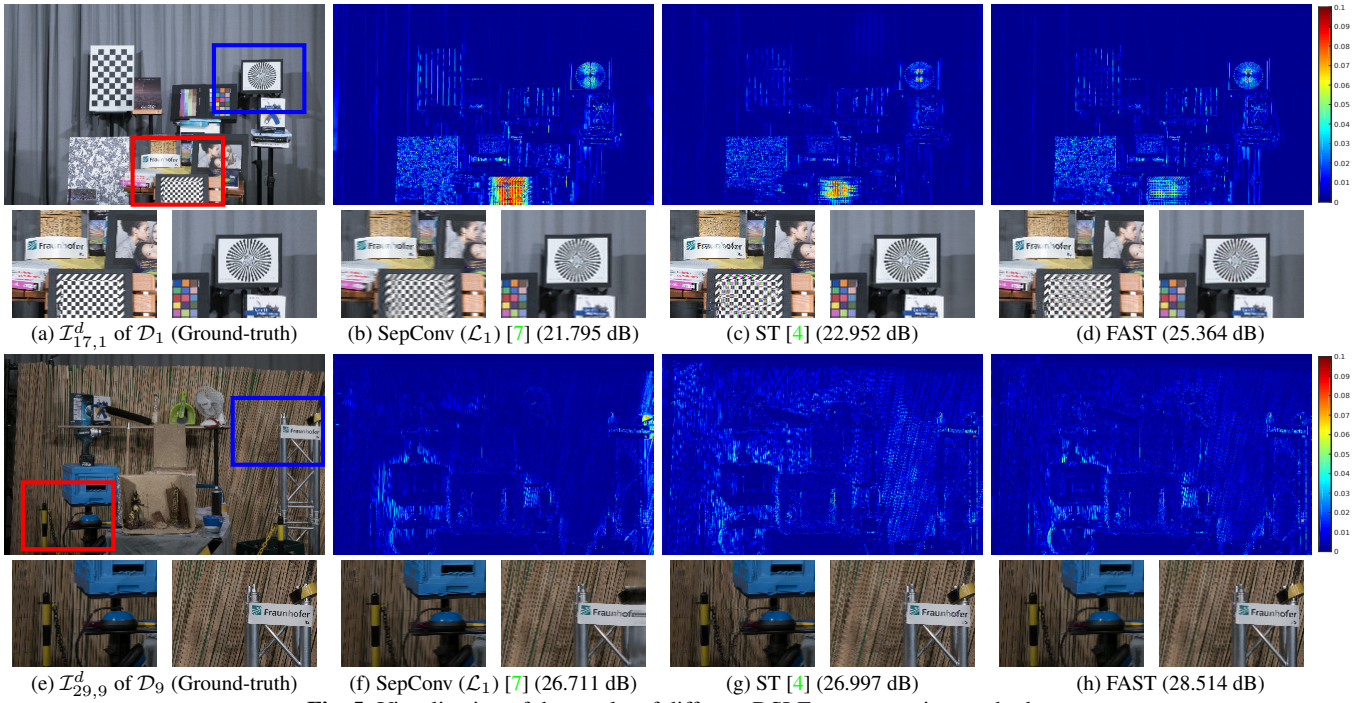
4.2. Results and analysis

The proposed method and other DSLF reconstruction approaches are evaluated quantitatively and qualitatively as follows:

Quantitative evaluation. The minimum per-view PSNR values of the DSLF reconstruction on all the nine evaluation SSLF sub-datasets $\mathcal{S}_\mu$ using different methods are exhibited in Table I. As can be seen from this table, the proposed FAST method achieves the best performance on most of the evaluation SSLF sub-datasets except for $\mathcal{S}_\mu$, $\mu \in \{4, 5\}$. However, on these two SSLF sub-datasets, the minimum per-view PSNR values of the FAST method are only 0.276 and 0.093 dB less than those of the ST approach. Besides, on

Table I. Minimum per-view PSNR values (in dB, explained in Sect. 4.1) for the performance evaluation of different DSLF reconstruction methods on evaluation SSLF sub-datasets $\mathcal{S}_\mu$ ($1 \leqslant \mu \leqslant 9$).

Method	$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$	$\mathcal{S}_4$	$\mathcal{S}_5$	$\mathcal{S}_6$	$\mathcal{S}_7$	$\mathcal{S}_8$	$\mathcal{S}_9$
SepConv ($\mathcal{L}_1$) [7]	21.018	17.579	20.330	22.215	24.955	25.924	18.141	18.781	22.715
PIASC ($\mathcal{L}_1$) [9]	21.015	17.572	20.332	22.221	24.961	25.929	18.140	18.784	22.709
ST [4]	22.699	17.634	20.231	**23.842**	**25.752**	26.527	18.015	18.727	22.304
FAST	**24.683**	**17.988**	**20.828**	23.566	25.659	**27.002**	**18.393**	**19.171**	**22.884**



(a) $\mathcal{I}^d_{17,1}$ of $\mathcal{D}_1$ (Ground-truth) (b) SepConv ($\mathcal{L}_1$) [7] (21.795 dB) (c) ST [4] (22.952 dB) (d) FAST (25.364 dB)

(e) $\mathcal{I}^d_{29,9}$ of $\mathcal{D}_9$ (Ground-truth) (f) SepConv ($\mathcal{L}_1$) [7] (26.711 dB) (g) ST [4] (26.997 dB) (h) FAST (28.514 dB)

Fig. 5. Visualization of the results of different DSLF reconstruction methods.

$\mathcal{S}_1$, the DSLF reconstruction results indicate that the FAST method achieves a significant improvement in performance over the second-best approach with a gain of 1.984 dB, which demonstrates the effectiveness of the proposed method. Moreover, it can also be seen from the data in this table that SepConv and PIASC behave almost the same and both of them outperform the ST approach on $\mathcal{S}_\mu$, $\mu \in \{3, 7, 8, 9\}$.

Qualitative evaluation. Since the above analysis shows the similar DSLF reconstruction performance results of SepConv and PIASC, only SepConv is visually evaluated here. Some of the reconstructed views from $\mathcal{S}_\mu$ using different methods are displayed in Fig. 5. The top row of this figure illustrates the DSLF reconstruction results *w.r.t.* $\mathcal{I}^d_{17,1}$ of $\mathcal{D}_1$. The image patches containing the checkerboard and Siemens star are selected as the interesting areas. The SepConv method totally fails in reconstructing the checkerboard part with blur and artifacts. In addition, the Siemens star reconstructed by it is also blurry. This is because that the size of the repetitive pattern, *i.e.* the checker, is smaller than the disparities of neighboring views of $\mathcal{S}_\mu$, which confuses the network of SepConv at the aspect of the real-motion decision. The ST approach does not have such motion-decision problem; however, it still outputs artifacts when reconstructing the checkerboard. The proposed FAST method alleviates this problem with visually appealing results for both interesting areas, partially because it exploits the disparity information. The bottom row of Fig. 5 shows the visualization results *w.r.t.* $\mathcal{I}^d_{29,9}$ of $\mathcal{D}_9$. Two image patches involving the toolbox and Fraunhofer logo with a background of bamboo curtain are chosen as the interesting areas. Looking at Fig. 5 (f), the SepConv method succeeds in reconstructing the toolbox part; however, it fails in the reconstruction of the border of the Fraunhofer-logo patch. The ST approach solves this blur-border problem, while it produces artifacts *w.r.t.* the bamboo curtain in the interesting area including the toolbox. The proposed FAST method successfully addresses these issues with achieving sharp and visually correct results for both interesting areas, which implies that FAST is effective in DSLF reconstruction for the scene having a complex background.

5. CONCLUSION

This paper presents a novel learning-based method, Flow-Assisted Shearlet Transform (FAST), for solving the DSLF reconstruction problem. The FAST method fully leverages a state-of-the-art optical flow method, *i.e.* PWC-Net, to estimate bidirectional disparity maps between adjacent views in an input SSLF, which are beneficial to the estimation of the inverse disparity maps and the preparation of the shearlet system in ST. Besides, FAST employs two fully convolutional neural networks, *i.e.* disparity refinement network and view synthesis network, to reconstruct a DSLF from a SSLF using the disparity information and ST results. Experimental results on nine challenging real-world DSLF sub-datasets with large disparity ranges show that the proposed FAST method achieves better DSLF reconstruction results than the other state-of-the-art approaches.

Acknowledgments. The work in this paper was funded from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 676401, European Training Network on Full Parallax Imaging, and the German Research Foundation (DFG) No. K02044/8-1. The Titan Xp used for this research was donated by the NVIDIA Corporation.

6. REFERENCES

[1] M. Levoy and P. Hanrahan, "Light field rendering," in *SIGGRAPH*, 1996, pp. 31–42. 1

[2] S. Vagharshakyan, R. Bregovic, and A. Gotchev, "Image based rendering technique via sparse representation in shearlet domain," in *ICIP*, 2015, pp. 1379–1383. 1

[3] S. Vagharshakyan, R. Bregovic, and A. Gotchev, "Light field reconstruction using shearlet transform," *IEEE TPAMI*, vol. 40, no. 1, pp. 133–147, 2018. 1, 2

[4] S. Vagharshakyan, R. Bregovic, and A. Gotchev, "Accelerated shearlet-domain light field reconstruction," *IEEE J-STSP*, vol. 11, no. 7, pp. 1082–1091, 2017. 1, 2, 3, 4

[5] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, "PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume," in *CVPR*, 2018, pp. 8934–8943. 1, 2

[6] S. Niklaus, L. Mai, and F. Liu, "Video frame interpolation via adaptive convolution," in *CVPR*, 2017, pp. 2270–2279. 1

[7] S. Niklaus, L. Mai, and F. Liu, "Video frame interpolation via adaptive separable convolution," in *ICCV*, 2017, pp. 261–270. 1, 4

[8] H. Jiang, D. Sun, V. Jampani, M.-H. Yang, E. Learned-Miller, and J. Kautz, "Super SloMo: High quality estimation of multiple intermediate frames for video interpolation," in *CVPR*, 2018, pp. 9000–9008. 1, 2

[9] Y. Gao and R. Koch, "Parallax view generation for static scenes using parallax-interpolation adaptive separable convolution," in *ICME Workshops*, 2018, pp. 1–4. 1, 3, 4

[10] Y. Gao, R. Bregovic, A. Gotchev, and R. Koch, "MAST: Mask-accelerated shearlet transform for densely-sampled light field reconstruction," in *ICME*, 2019. 1

[11] M. Genzel and G. Kutyniok, "Asymptotic analysis of inpainting via universal shearlet systems," *SIAM Journal on Imaging Sciences*, vol. 7, no. 4, pp. 2301–2339, 2014. 2

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015, pp. 234–241. 2

[13] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldluecke, "A dataset and evaluation methodology for depth estimation on 4D light fields," in *ACCV*, 2016, pp. 19–34. 2

[14] M. Ziegler, R. op het Veld, J. Keinert, and F. Zilly, "Acquisition system for dense lightfield of large scenes," in *3DTV-CON*, 2017, pp. 1–4. 3