IEST: Interpolation-Enhanced Shearlet Transform for Light Field Reconstruction Using Adaptive Separable Convolution

Yuan Gao, Reinhard Koch
Department of Computer Science
Kiel University
24118 Kiel, Germany
{yga, rk}@informatik.uni-kiel.de

Robert Bregovic, Atanas Gotchev
Faculty of Information Technology and Communication Sciences
Tampere University
33014 Tampere, Finland
{robert.bregovic, atanas.gotchev}@tuni.fi

***Abstract*—The performance of a light field reconstruction algorithm is typically affected by the disparity range of the input Sparsely-Sampled Light Field (SSLF). This paper finds that (i) one of the state-of-the-art video frame interpolation methods, *i.e.* adaptive Separable Convolution (SepConv), is especially effective for the light field reconstruction on a SSLF with a small disparity range (< 10 pixels); (ii) one of the state-of-the-art light field reconstruction methods, *i.e.* Shearlet Transformation (ST), is especially effective in reconstructing a light field from a SSLF with a moderate disparity range (10-20 pixels) or a large disparity range (> 20 pixels). Therefore, to make full use of both methods to solve the challenging light field reconstruction problem on SSLFs with moderate and large disparity ranges, a novel method, referred to as Interpolation-Enhanced Shearlet Transform (IEST), is proposed by incorporating these two approaches in a coarse-to-fine manner. Specifically, ST is employed to give a coarse estimation for the target light field, which is then refined by SepConv to improve the reconstruction quality of parallax views involving small disparity ranges. Experimental results show that IEST outperforms the other state-of-the-art light field reconstruction methods on nine challenging horizontal-parallax evaluation SSLF datasets of different real-world scenes with moderate and large disparity ranges.**

***Index Terms*—Light Field Reconstruction, Parallax View Generation, Adaptive Separable Convolution, Shearlet Transform, Interpolation-Enhanced Shearlet Transform**

I. Introduction

A light field can be considered as a 4D approximation of the plenoptic function parameterized by two parallel planes (camera plane and image plane) [1]; therefore, a 4D light field is typically composed of camera images sampled on a regular 2D grid [2] or an irregular 2D grid [3]. If the disparities between adjacent views in a light field are less than one pixel, this light field can be referred to as a Densely-Sampled Light Field (DSLF) [4]. How to capture a horizontal-parallax light field is illustrated in Fig. 1. As can be seen from this figure, the horizontal-parallax desired light field is captured by a system with cameras uniformly distributed along the horizontal axis 's' with the same camera orientation. Let this desired target light field be denoted by $\mathcal{D} = \{\mathcal{I}_i | 1 \leqslant i \leqslant m\}$. Due to the hardware limitations of most of the light field capture systems in real-world environments, it is difficult for them to capture all the m parallax images of the desired target light field $\mathcal{D}$ with small disparities. In other words, such camera systems can capture only part of parallax views of the target light field $\mathcal{D}$, which are represented by the solid-line green cameras in Fig. 1. Let this Sparsely-Sampled Light Field (SSLF) be denoted by $\mathcal{S}$, $\mathcal{S} \subseteq \mathcal{D}$ and $|\mathcal{S}| = n$ $(< m)$. This paper aims to solve the problem of reconstructing the missing parallax views for the target light field $\mathcal{D}$ from the input SSLF $\mathcal{S}$. The relationship between them is determined by the interpolation rate δ, where $\delta = \frac{m-1}{n-1}$. It is obvious that different interpolation rates correspond to different disparity conditions for the input SSLF $\mathcal{S}$. Besides, if the target light field $\mathcal{D}$ is a DSLF, the light field reconstruction on $\mathcal{S}$ can be called DSLF reconstruction.

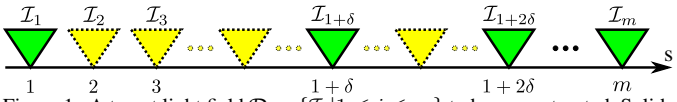


Figure 1. A target light field $\mathcal{D} = \{\mathcal{I}_i | 1 \leqslant i \leqslant m\}$ to be reconstructed. Solid-line green triangles constitute an input SSLF $\mathcal{S}$. Dash-line yellow triangles are the missing parallax views to be reconstructed for the target light field $\mathcal{D}$.

Motivation. The adaptive Separable Convolution (SepConv) approach [5] is one of the state-of-the-art video frame interpolation methods, which is extended by Gao and Koch in [6] for solving the DSLF reconstruction problem in a recursive manner, treating an input horizontal-parallax light field as a video captured by a virtual camera moving horizontally. The restriction of SepConv is that it may fail in light field reconstruction on SSLFs with larger disparity ranges, because its novel view synthesis ability is limited by the size of the convolution kernels. For more details refer to Sect. IV-B. Alternatively, one of the state-of-the-art light field reconstruction methods, *i.e.* Shearlet Transform (ST) [7, 8], is a universal solution to DSLF reconstruction and does not suffer from such restriction. This paper focuses on investigating how to employ the advantages of these two methods to better reconstruct light fields from SSLFs with moderate and large disparity ranges.

To address the challenging light field reconstruction problem for the cases of moderate and large disparity ranges, a novel method, referred to as Interpolation-Enhanced Shearlet Transform (IEST), is proposed in this paper. The proposed IEST method fully leverages the advantages of both ST and

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

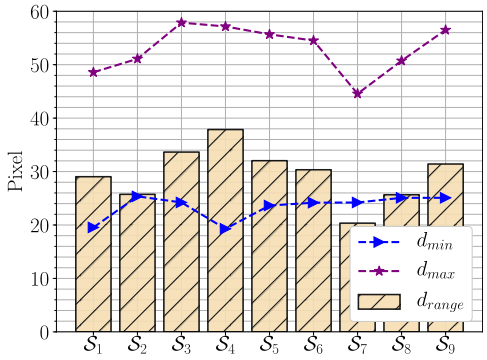
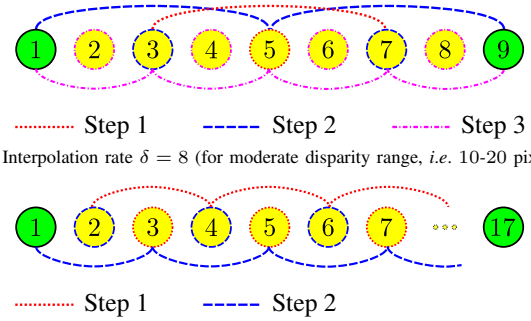


Figure 2. Disparity estimation for nine evaluation SSLFs $\mathcal{S}_\mu$ with an interpolation rate $\delta = 16$ using PWC-Net [9].



(a) Interpolation rate $\delta = 8$ (for moderate disparity range, *i.e.* 10-20 pixels)

(b) Interpolation rate $\delta = 16$ (for large disparity range, *i.e.* > 20 pixels)

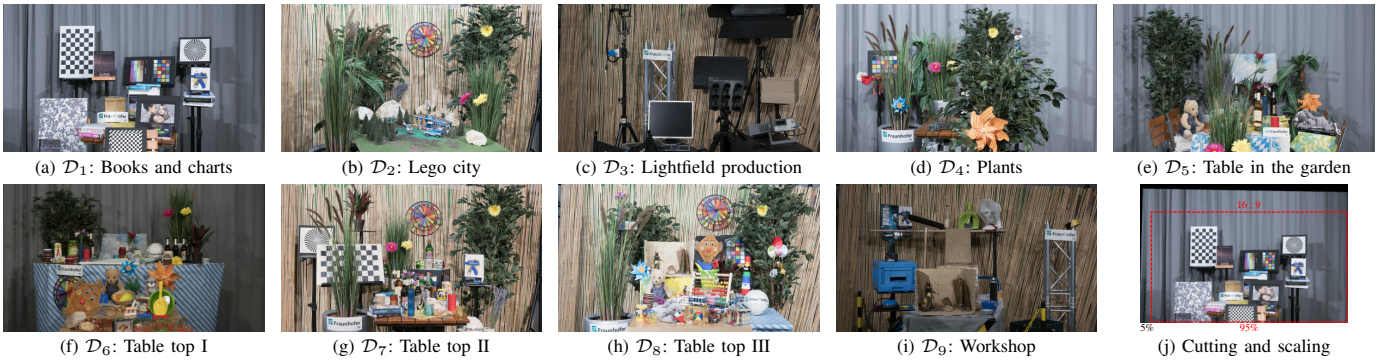
Figure 3. Flowcharts of IEST for light field reconstruction from SSLFs at different interpolation rates, *i.e.* $\delta \in \{8, 16\}$.

SepConv in a coarse-to-fine manner to reconstruct a target light field from a horizontal-parallax SSLF with a moderate or large disparity range. Specifically, ST is applied to reconstruct the target light field $\mathcal{D}$ from an input SSLF $\mathcal{S}$, so that the missing parallax views in $\mathcal{S}$ are coarsely estimated. Two elaborately-designed parallax view refinement strategies, corresponding to different interpolation rates $\delta \in \{8, 16\}$, are then performed on the coarsely-estimated $\mathcal{D}$ in a recursive way. Experimental results indicate that IEST outperforms all the other state-of-the-art methods on nine challenging horizontal-parallax evaluation SSLF datasets for both the moderate disparity range (10-19 pixels) and the large disparity range (20-38 pixels). Moreover, for any evaluation SSLF dataset with a small disparity range (5-9.5 pixels), SepConv achieves better light field reconstruction performance than ST.

II. Related Work

Learning-based video frame synthesis. Niklaus *et al*. employ a deep fully Convolutional Neural Network (CNN) to estimate pixel-wise spatially-adaptive 2D convolution kernels, which are applied on the two consecutive input video frames to synthesize an intermediate one [10]. However, for each image pixel, this method predicts a $n \times n$ ($n = 41$) convolution kernel, which will be prohibitively expensive in memory requirement if the input images are in high resolution. To tackle this problem, Niklaus *et al*. propose a spatially-adaptive Separable Convolution (SepConv) approach, which approximates each of the 2D convolution kernels with a pair of 1D kernels, thus reducing the number of kernel parameters from n^2 to $2n$ for each 2D convolution kernel [5]. Liu *et al*. propose an end-to-end deep network, *i.e*. Deep Voxel Flow (DVF), to synthesize a video frame in either interpolation or extrapolation with sharp results [11]. More recently, Niklaus *et al*. fully leverage a state-of-the-art optical flow algorithm, *i.e*. PWC-Net [9], to estimate bidirectional flow between two consecutive input video frames, which is applied to pre-warp the input video frames together with their corresponding per-pixel context maps extracted by a pre-trained neural network [12]. All these pieces of pre-warped information are then fed to a video frame synthesis network, *i.e*. a modified GridNet [13], to interpolate an intermediate video frame at a desired temporal position. Jiang *et al*. also estimate bidirectional optical flow between two consecutive input video frames via a flow computation CNN [14]. The estimated optical flow is then refined by a flow interpolation CNN, which additionally predicts a soft visibility map. Both the refined optical flow and predicted soft visibility map are utilized to interpolate an intermediate video frame at any arbitrary time step via warping and fusion. Meyer *et al*. apply the steerable pyramid filters [15] to decompose the input two consecutive video frames [16]. Their decompositions, consisting of amplitudes, phases and low-pass residuals, are fed to a decoder-only neural network, *i.e*. PhaseNet, to predict the corresponding decomposition of the intermediate frame in order to fulfill image reconstruction. Visually preferable results are achieved by this method in challenging scenarios containing lighting changes or motion blur.

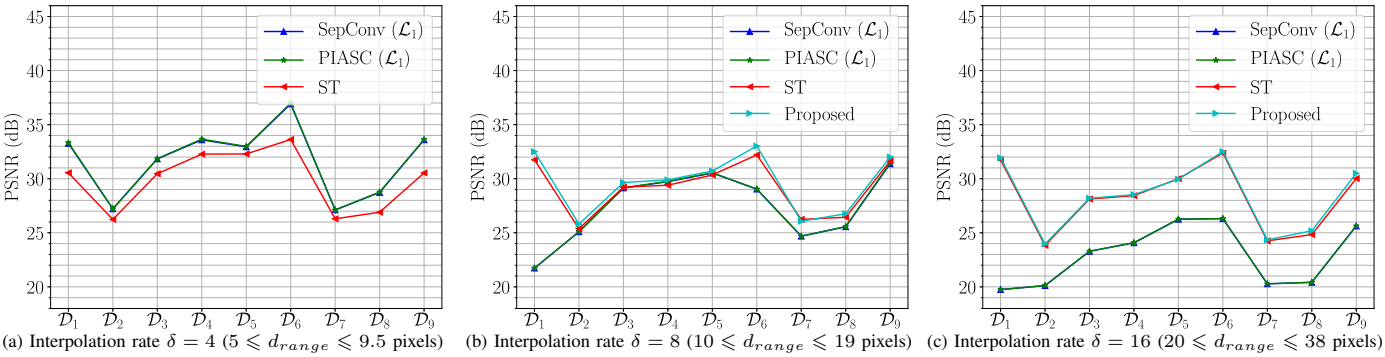
Light field angular super-resolution. Kalantari *et al*. propose a learning-based view synthesis approach, which is composed of disparity and color estimators, for synthesizing novel views from a sparse set of sub-aperture images of a micro-lens array-based consumer light field camera [17]. Wu *et al*. present a blur-restoration-deblur framework for Epipolar-Plane Image (EPI) interpolation to reconstruct dense light fields [18]. A residual network with three convolution layers is utilized to restore the angular detail of a blurred and up-sampled EPI. However, due to the limitation in the blurring kernel size and bicubic interpolation, this method can only handle SSLF data with very small disparity ranges (up to 5 pixels). Vagharshakyan *et al*. reconstruct DSLFs from SSLFs by exploiting EPI sparsification in shearlet domain, which has been demonstrated to be effective in reconstructing Lambertian scenes and non-Lambertian scenes containing semi-transparent objects [7, 8]. Gao and Koch employ a fine-tuning strategy to enhance the motion-sensible convolution kernels of the state-of-the-art video frame interpolation method, *i.e*. SepConv, and propose Parallax-Interpolation Adaptive Separable Convolution (PIASC) to reconstruct a DSLF from a horizontal-parallax SSLF in a recursive way [6]. Yeung *et al*. design an end-to-end 4D convolutional light field reconstruction network consisting of view synthesis and view refinement phases for fast light field reconstruction from a SSLF [19]. Wang *et al*. also propose an end-to-end learning framework for fast light field reconstruction [20]. Their network includes two 2D strided convolutions for the interpolation of stacked sparsely-sampled EPIs and two detail-restoration 3D CNNs for restoring high-frequency details of these interpolated EPI volumes. In conclusion, studies in [19, 20] can only be applied on full-



(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

Figure 4. The middle views of nine evaluation datasets. Sub-image (j) illustrates the image cutting and scaling strategy in Sect. IV-A.



(a) Interpolation rate $\delta = 4$ ($5 \leqslant d_{range} \leqslant 9.5$ pixels) (b) Interpolation rate $\delta = 8$ ($10 \leqslant d_{range} \leqslant 19$ pixels) (c) Interpolation rate $\delta = 16$ ($20 \leqslant d_{range} \leqslant 38$ pixels)

Figure 5. Minimum per-view PSNR results (in dB, explained in Sect. IV-A) of different light field reconstruction methods on nine evaluation datasets with different interpolation rates $\delta \in \{4, 8, 16\}$.

parallax SSLF data. In addition, for different intermediate-view interpolation factors, *i.e.* δ in Sect. I, these methods need to re-train their networks. Nevertheless, this paper focuses on investigating a universal light field reconstruction solution *w.r.t.* input SSLFs with moderate and large disparity ranges.

III. Methodology

A. Shearlet Transform (ST)

For tackling the DSLF reconstruction problem on SSLFs with varying disparities, ST is originally proposed in [7] and extended in [8, 21, 22]. The key idea of ST is to design an elaborately-tailored universal shearlet system [7, 23], which is exploited to perform sparsity regularization in the shearlet transform domain for the sparsely-sampled EPIs of an input SSLF via an iterative α-adaptive algorithm [7] or a double overrelaxation (DORE) algorithm [8]. The performance of the ST algorithm relies on the precision of the disparity estimation of the input SSLF. Specifically, the minimum disparity d_{min} and maximum disparity d_{max} of this SSLF should be precisely estimated before applying ST. The corresponding disparity range of the input SSLF is derived from d_{min} and d_{max}, *i.e.* $d_{range} = (d_{max} - d_{min})$. Based on the value of the estimated d_{min}, a pre-shearing step using cubic interpolation is then applied on the input SSLF, so that the new minimum disparity $d'_{min} = 0$, the new maximum disparity $d'_{max} = d_{range}$ and the sheared input SSLF is able to be effectively and efficiently processed by a shearlet system with ξ scales, where $\xi = \lceil \log_2 d_{range} \rceil$. Finally, a post-processing shearing procedure is performed on the reconstructed DSLF in order to compensate for the loss of the minimum disparity that is eliminated in the pre-shearing step.

B. Interpolation-Enhanced Shearlet Transform (IEST)

Although ST is a universal solution to the light field reconstruction problem on SSLFs with varying disparities, it is not as effective as one of the state-of-the-art video frame interpolation methods, *i.e.* SepConv, for light field reconstruction from SSLFs with small disparity ranges. An example for this phenomenon is shown in Fig. 5 (a), where the interpolation rate $\delta = 4$ equals to $5 \leqslant d_{range} \leqslant 9.5$ pixels, derived from $20 \leqslant d_{range} \leqslant 38$ pixels in the case of $\delta = 16$ in Fig. 2, for all the evaluation SSLFs $\mathcal{S}_\mu$. However, for SSLFs with moderate and large disparity ranges, ST tends to be more effective than SepConv as illustrated in Fig. 5 (b) and (c). Intuitively, taking advantage of SepConv to refine the parallax views of light fields that are reconstructed by ST from SSLFs with moderate and large disparity ranges may improve the final light field reconstruction performance. Therefore, a novel light field reconstruction method, *i.e.* Interpolation-Enhanced Shearlet Transform (IEST), is proposed. The IEST method is specially designed for light field reconstruction on SSLFs with moderate and large disparity ranges with a consideration that the reconstructed parallax views of ST involving small disparity ranges can be refined by SepConv. Depending on different interpolation rates, two parallax view refinement strategies of IEST are presented in Fig. 3. As shown in (a), the first strategy is designed for the case of interpolation rate $\delta = 8$. Here, green circles stand for the ground-truth parallax views from an input SSLF $\mathcal{S}$ (also see Fig. 1) and yellow circles denote the parallax views reconstructed by ST. The reconstructed parallax views represented by yellow circles are then refined by SepConv recursively, which is depicted by using three types of dash lines that represent Step 1, 2 and 3.

Table I
MINIMUM AND AVERAGE PER-VIEW PSNR RESULTS (IN DB, EXPLAINED IN SECT. IV-A) FOR THE PERFORMANCE EVALUATION OF DIFFERENT LIGHT FIELD RECONSTRUCTION METHODS ON NINE EVALUATION DATASETS.

Minimum PSNR
(Interpolation rate $\delta = 8$ and $10 \leqslant d_{range} \leqslant 19$ pixels)

	SepConv ($\mathcal{L}_1$) [5]	PIASC ($\mathcal{L}_1$) [6]	ST [8]	Proposed
$\mathcal{D}_1$	21.733	21.731	31.750	**32.505**
$\mathcal{D}_2$	25.087	25.103	25.375	**25.807**
$\mathcal{D}_3$	29.145	29.161	29.220	**29.644**
$\mathcal{D}_4$	29.729	29.760	29.399	**29.893**
$\mathcal{D}_5$	30.525	30.557	30.349	**30.713**
$\mathcal{D}_6$	29.039	29.044	32.203	**33.023**
$\mathcal{D}_7$	24.685	24.688	**26.237**	26.067
$\mathcal{D}_8$	25.558	25.576	26.438	**26.763**
$\mathcal{D}_9$	31.379	31.466	31.644	**32.005**

Minimum PSNR
(Interpolation rate $\delta = 16$ and $20 \leqslant d_{range} \leqslant 38$ pixels)

	SepConv ($\mathcal{L}_1$) [5]	PIASC ($\mathcal{L}_1$) [6]	ST [8]	Proposed
$\mathcal{D}_1$	19.753	19.742	31.754	**31.941**
$\mathcal{D}_2$	20.118	20.123	23.839	**23.964**
$\mathcal{D}_3$	23.289	23.295	28.125	**28.194**
$\mathcal{D}_4$	24.073	24.084	28.430	**28.529**
$\mathcal{D}_5$	26.254	26.262	**30.004**	29.955
$\mathcal{D}_6$	26.307	26.317	32.368	**32.504**
$\mathcal{D}_7$	20.283	20.283	24.257	**24.350**
$\mathcal{D}_8$	20.419	20.423	24.831	**25.184**
$\mathcal{D}_9$	25.628	25.639	30.021	**30.498**

Average PSNR
(Interpolation rate $\delta = 8$ and $10 \leqslant d_{range} \leqslant 19$ pixels)

	SepConv ($\mathcal{L}_1$) [5]	PIASC ($\mathcal{L}_1$) [6]	ST [8]	Proposed
$\mathcal{D}_1$	24.498	24.495	33.429	**33.826**
$\mathcal{D}_2$	26.625	26.648	26.666	**27.144**
$\mathcal{D}_3$	30.781	30.828	30.875	**31.534**
$\mathcal{D}_4$	32.344	32.446	32.189	**32.972**
$\mathcal{D}_5$	31.762	31.840	32.328	**32.613**
$\mathcal{D}_6$	31.333	31.383	34.843	**36.034**
$\mathcal{D}_7$	27.209	27.228	27.383	**27.827**
$\mathcal{D}_8$	27.299	27.332	27.732	**28.151**
$\mathcal{D}_9$	32.878	32.951	32.836	**33.407**

Average PSNR
(Interpolation rate $\delta = 16$ and $20 \leqslant d_{range} \leqslant 38$ pixels)

	SepConv ($\mathcal{L}_1$) [5]	PIASC ($\mathcal{L}_1$) [6]	ST [8]	Proposed
$\mathcal{D}_1$	21.514	21.505	33.220	**33.435**
$\mathcal{D}_2$	22.779	22.807	25.285	**25.467**
$\mathcal{D}_3$	26.009	26.044	29.764	**29.833**
$\mathcal{D}_4$	27.408	27.483	30.677	**30.933**
$\mathcal{D}_5$	28.005	28.063	31.337	**31.457**
$\mathcal{D}_6$	28.779	28.838	34.100	**34.224**
$\mathcal{D}_7$	23.375	23.397	25.915	**25.992**
$\mathcal{D}_8$	23.158	23.196	26.542	**26.775**
$\mathcal{D}_9$	28.416	28.478	32.083	**32.328**

To be precise, each step uses two parallax views having a small disparity range to synthesize the middle view between them. For the interpolation rate $\delta = 16$, the second strategy of IEST, as shown in (b), refines the reconstructed parallax views from ST with only two steps. It is worth to be mentioned that the strategy of IEST designed for $\delta = 8$ is especially effective for the light field reconstruction on SSLFs with moderate disparity ranges (10-20 pixels), while the second IEST strategy designed for $\delta = 16$ is more effective for the light field reconstruction on SSLFs with large disparity ranges (> 20 pixels).

IV. EXPERIMENTS

A. Experimental Settings

Evaluation datasets. The dataset of High Density Camera Array (HDCA) [24] is used for evaluating the performance of all methods. This dataset has nine high-fidelity dense light fields for different real-world scenes captured by a movable high-resolution and high-quality DSLR camera. Eight of them have the same angular resolution of 101×21. The remaining one has an angular resolution of 99×21. The spatial resolution of these light fields is 3976×2652 pixels. Since the proposed method and baseline approaches are originally designed for parallax view generation using horizontal-parallax SSLFs, only the top 97 horizontal-parallax views of each light field are selected for evaluation. However, these raw images have a problem that some boundary regions have no color information due to calibration, which is not fair for performance comparison of different methods. To overcome this limitation, an image cutting and scaling strategy is proposed as illustrated Fig. 4 (j). In particular, a 95%-width image (at the right of the original raw image) is cut and a $16:9$-shape image at the bottom of this cut image is then downscaled to a new resolution of 1280×720 pixels using bicubic interpolation. After performing these two operations for all the light fields, nine horizontal-parallax light field datasets $\mathcal{D}_\mu$ are constructed and their middle views are exhibited in Fig. 4 (a)-(i). Note that $m = 97$ for each ground-truth light field dataset $\mathcal{D}_\mu$, where $1 \leqslant \mu \leqslant 9$.

Disparity estimation. From these nine ground-truth light field datasets $\mathcal{D}_\mu$, the corresponding input SSLFs S_μ are constructed by using different interpolation rates δ, *i.e.* $\delta \in \{4, 8, 16\}$, as introduced in Sect. I. In order to perform ST method correctly, the disparity conditions of different SSLFs $\mathcal{S}_\mu$ should be estimated precisely. To tackle this problem, a state-of-the-art optical flow method, *i.e.* PWC-Net [9], is applied to estimate the bidirectional flow between neighboring views in $\mathcal{S}_\mu$ for the case of $\delta = 16$. Note that only the horizontal components of the calculated optical flow contain useful information, which are leveraged to compute d_{min}, d_{max} and d_{range} for each $\mathcal{S}_\mu$ as illustrated in Fig. 2. It can be found that the minimum d_{range} for all the SSLFs $\mathcal{S}_\mu$ with the interpolation rate of 16 is around 20 pixels, which suggests that all the target light fields $\mathcal{D}_\mu$ are not DSLFs.

Evaluation criteria. The per-view PSNR is exploited to evaluate the performance of different light field reconstruction methods. Additionally, for any parallax view generation method evaluated on a dataset $\mathcal{D}_\mu$, the minimum and average per-view PSNR values constitute the final evaluation criteria.

Implementation details. All the methods mentioned in this paper are implemented using CUDA and executed on an NVIDIA Titan Xp GPU. The pre-trained neural networks of SepConv and PWC-Net are from [5] and [9], respectively. Besides, the parameters of ST using the DORE algorithm are set as same as [8], where $\alpha = 20$ with 100 iterations and a low-pass initial estimation. The construction of the shearlet system used by ST relies on the estimated disparity range of the input SSLF S_μ, as explained in Sect. III-A.

B. Results and Analysis

The minimum per-view PSNR values using different light field reconstruction methods on all the evaluation SSLFs $\mathcal{S}_\mu$ at varying interpolation rates, *i.e.* $\delta \in \{4, 8, 16\}$, are presented in Fig. 5. Comparing (a), (b) and (c) in this figure, it can be seen that SepConv achieves better performance than ST on all the evaluation DSLFs for the case that interpolation rate

$\delta = 4$. However, for higher interpolation rates, *i.e.* $\delta \in \{8, 16\}$, the performance of SepConv is significantly worse than that of ST for reconstructing the target light fields $\mathcal{D}_\mu$ from $\mathcal{S}_\mu$. The main reason for this is that (i) SepConv is not capable of correctly interpolating novel views with repetitive patterns that are smaller than the disparities between the input neighboring parallax views, *e.g.*, checkers of the checkerboards in Fig. 4 (a); (ii) the size of the convolution kernel of SepConv is 51×51 pixels, which restricts its novel view synthesis ability *w.r.t.* two parallax images with moderate or large disparities. Besides, directly increasing the convolutional kernel size of SepConv involves re-training the whole network of SepConv and increasing the memory demand, which will not be the best solution to the light field reconstruction problem.

The minimum and average per-view PSNR values of different light field reconstruction methods for interpolation rates $\delta \in \{8, 16\}$ are also shown in Table I. The top row of this table presents the light field reconstruction results for the case of $\delta = 8$, corresponding to $10 \leqslant d_{range} \leqslant 19$ pixels. It can be found that (i) for the minimum-PSNR evaluation criteria, the proposed IEST method achieves the best performance on most of the evaluation DSLFs except for $\mathcal{D}_7$; (ii) for the average-PSNR evaluation criteria, IEST performs significantly better than all the baseline approaches. In addition, on $\mathcal{D}_6$, IEST yields a substantial performance gain of 0.82 and 1.191 dB *w.r.t.* minimum and average PSNRs over the second-best method, *i.e.* ST. This indicates that the proposed IEST method is effective in light field reconstruction on SSLFs with moderate disparity ranges, *e.g.*, up to 19 pixels in given examples. Moreover, SepConv and PIASC have almost the same performance, which is better than ST on $\mathcal{D}_\mu$, $\mu \in \{4, 5\}$ *w.r.t.* minimum-PSNR criterion. The bottom row of Table I shows the light field reconstruction results for the case of $\delta = 16$, corresponding to $20 \leqslant d_{range} \leqslant 38$ pixels. The proposed method outperforms all the other baseline methods on all the evaluation DSLFs except for $\mathcal{D}_5$, where the minimum PSNR of IEST is only 0.049 dB less than that of ST. This suggests that the proposed IEST method is also effective for reconstructing light fields from SSLFs with large disparity ranges, *e.g.*, up to 38 pixels in given examples.

V. Conclusion

In this paper, a novel light reconstruction method, *i.e.* IEST, is presented for reconstructing target light fields from input SSLFs with moderate and large disparity ranges. IEST takes full advantage of a state-of-the-art DSLF reconstruction method, *i.e.* ST, and a state-of-the-art video frame interpolation method, *i.e.* SepConv, to perform light field angular super-resolution in a coarse-to-fine manner. Specifically, SepConv is utilized to refine the light field reconstruction results of ST in a recursive way. Experimental results on nine challenging evaluation datasets demonstrate the effectiveness of IEST over the other state-of-the-art light field reconstruction approaches for both the moderate disparity range (10-19 pixels) and the large disparity range (20-38 pixels).

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.

References

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

[2] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," in *ACM TOG*, 2005, vol. 24, pp. 765–776.

[3] B. Mildenhall, P. P. Srinivasan, R. Ortiz-Cayon, N. K. Kalantari, R. Ramamoorthi, R. Ng, and A. Kar, "Local light field fusion: Practical view synthesis with prescriptive sampling guidelines," *ACM TOG*, vol. 38, no. 4, pp. 29:1–29:14, 2019.

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

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

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

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

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

[9] 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.

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

[11] Z. Liu, R.A. Yeh, X. Tang, Y. Liu, and A. Agarwala, "Video frame synthesis using deep voxel flow," in *ICCV*, 2017, pp. 4473–4481.

[12] S. Niklaus and F. Liu, "Context-aware synthesis for video frame interpolation," in *CVPR*, 2018, pp. 1701–1710.

[13] D. Fourure, R. Emonet, E. Fromont, D. Muselet, A. Trémeau, and C. Wolf, "Residual Conv-Deconv grid network for semantic segmentation," in *BMVC*, 2017.

[14] 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.

[15] E.P. Simoncelli and W.T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *ICIP*, 1995, vol. 3, pp. 444–447.

[16] S. Meyer, A. Djelouah, B. McWilliams, A. Sorkine-Hornung, M. Gross, and C. Schroers, "PhaseNet for video frame interpolation," in *CVPR*, 2018, pp. 498–507.

[17] N. K. Kalantari, T.-C. Wang, and R. Ramamoorthi, "Learning-based view synthesis for light field cameras," *ACM TOG*, vol. 35, no. 6, pp. 193:1–193:10, 2016.

[18] G. Wu, M. Zhao, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field reconstruction using deep convolutional network on EPI," in *CVPR*, 2017, pp. 1638–1646.

[19] H.W.F. Yeung, J. Hou, J. Chen, Y.Y. Chung, and X. Chen, "Fast light field reconstruction with deep coarse-to-fine modelling of spatial-angular clues," in *ECCV*, 2018.

[20] Y. Wang, F. Liu, Z. Wang, G. Hou, Z. Sun, and T. Tan, "End-to-end view synthesis for light field imaging with pseudo 4DCNN," in *ECCV*, 2018, pp. 340–355.

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

[22] Y. Gao, R. Koch, R. Bregovic, and A. Gotchev, "FAST: Flow-assisted shearlet transform for densely-sampled light field reconstruction," in *ICIP*, 2019.

[23] 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.

[24] 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.