MAST: Mask-accelerated shearlet transform for densely-sampled light field reconstruction

Citation

Year
2019

Version
Peer reviewed version (post-print)

Link to publication
TUTCRIS Portal (http://www.tut.fi/tutcris)

Published in
2019 IEEE International Conference on Multimedia and Expo, ICME 2019

DOI
10.1109/ICME.2019.00040

Take down policy
If you believe that this document breaches copyright, please contact cris.tau@tuni.fi, and we will remove access to the work immediately and investigate your claim.
ABSTRACT
Shearlet Transform (ST) is one of the most effective algorithms for the Densely-Sampled Light Field (DSLF) reconstruction from a Sparsely-Sampled Light Field (SSLF) with a large disparity range. However, ST requires a precise estimation of the disparity range of the SSLF in order to design a shearlet system with decent scales and to pre-shear the sparsely-sampled Epipolar-Plane Images (EPIs) of the SSLF. To overcome this limitation, a novel coarse-to-fine DSLF reconstruction method, referred to as Mask-Accelerated Shearlet Transform (MAST), is proposed in this paper. Specifically, a state-of-the-art learning-based optical flow method, FlowNet2, is employed to estimate the disparities of a SSLF. The estimated disparities are then utilized to roughly estimate the densely-sampled EPIs for the sparsely-sampled EPIs of the SSLF. Finally, an elaborately-designed soft mask for a coarsely-inpainted EPI is exploited to perform an iterative refinement on this EPI. Experimental results on nine challenging horizontal-parallax real-world SSLF datasets with large disparity ranges (up to 35 pixels) demonstrate the effectiveness and efficiency of the proposed method over the other state-of-the-art approaches.

Index Terms—View Synthesis, Parallax View Generation, Densely-Sampled Light Field Reconstruction, Shearlet Transform, Mask-Accelerated 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]. As can be seen in Fig. 1 (a), a horizontal-parallax light field capture system can be considered as a camera moving along the horizontal axis. All the parallax views captured by this camera constitute a ground-truth 3D light field volume as illustrated in Fig. 1 (b). This volume can then be turned into ground-truth Epipolar-Plane Images (EPIs), of which an example is shown in Fig. 1 (c). A Sparsely-Sampled Light Field (SSLF) for this horizontal-parallax light field dataset consists of views with blue borders. The virtual cameras represented by dash-line triangles with yellow color correspond to the target "unknown" views to be reconstructed, the number of which is decided by the interpolation rate δ. A sparsely-sampled EPI ε from the ground-truth EPI ψ is presented in Fig. 1 (d). The DSLF reconstruction for the SSLF can be treated as reconstructing a densely-sampled EPI f from the sparsely-sampled EPI ε. If the ground-truth EPI ψ is not densely sampled, it will be necessary to down-sample the reconstructed densely-sampled EPI f to construct a target EPI with the same size as ψ (see Sect. 4.1).

Shearlet Transform (ST) [3, 4] is extremely effective in reconstructing a densely-sampled EPI from a sparsely-sampled EPI with a large disparity range. This algorithm typically needs to obtain the disparity range of the sparsely-sampled EPI to construct a specifically-tailored universal shearlet system [3, 5] with decent scales. Besides, the sparsely-sampled EPI also needs the disparity range information for shearing and padding in order to be correctly processed by this elaborately-designed shearlet system. Moreover, for DSLF reconstruction from SSLFs with large disparity ranges, this algorithm is prone to be time-consuming due to the high num-
ber of iterations of its iterative thresholding algorithm. Therefore, in this paper, a novel ST-based coarse-to-fine DSFL reconstruction method, referred to as Mask-Accelerated Shearlet Transform (MAST), is proposed to address these two problems. The presented MAST method takes full advantage of a state-of-the-art learning-based optical flow estimation approach, i.e., FlowNet2 [6], to estimate the disparities of the whole SSLF for resolving the first problem. In addition, the estimated disparities are also used to roughly restore a densely-sampled EPI from a sparsely-sampled EPI via inverse warping. The iterative estimation refinement algorithm in ST convergences faster by means of an elaborately-designed soft mask for the coarsely-inpainted densely-sampled EPI, thus tackling the second problem. Experimental results demonstrate the superior performance of MAST over the other state-of-the-art DSFL reconstruction methods on nine challenging horizontal-parallax real-world light field datasets with disparity ranges up to 35 pixels.

2. RELATED WORK
High-quality and high-fidelity Virtual Reality (VR) [7] and Free Viewpoint Video (FVV) [8] contents fundamentally rely on DSFLs for the reason that DSFLs can be turned into continuous light fields via linear interpolation [9]. However, due to the difficulty of directly capturing a DSFL, a DSFL is typically reconstructed from a SSLF. The challenging DSFL reconstruction problem has been tried to be solved by light field angular super-resolution-based approaches, most of which treat it as novel view synthesis problem and do not consider the disparity range of the input SSLF. Kalantari et al. propose a learning-based approach composed of a disparity estimator and a color predictor to synthesize novel views from four corner sub-aperture views of a micro-lens array-based light field camera [10]. Wu et al. utilize a residual-learning method to restore the angular detail of EPIs within a blur-deblur framework [11]. However, the maximum disparity of the SSLF that can be handled by this approach is only 5 pixels. Yeung et al. also design a learning-based view synthesis network consisting of view synthesis and refinement components to reconstruct DSFLs [12]. Nevertheless, for different interpolation rates, their network needs to be retrained. Gao and Koch utilize a state-of-the-art video frame interpolation method, i.e., adaptive Separable Convolution (SepConv) [13], and a fine-tuning strategy enhancing the convolution kernels of SepConv to reconstruct DSFLs in a recursive way [14].

3. METHODOLOGY
3.1. DSFL reconstruction using ST
The shearlet transform approach for DSFL reconstruction is originally proposed in [3] and extended in [4] with computational acceleration. Given a coarsely-sampled EPI \( \epsilon \in \mathbb{R}^{m \times n} \) from a SSLF as shown in Fig. 1 (d), ST reconstructs a desired densely-sample EPI \( f \in \mathbb{R}^{m \times n} \) by an iterative inpainting algorithm using the sparse representation of \( f \) in shearlet domain. The sampling interval of the desired EPI \( f \) for rearranging the rows of the input decimated EPI \( \epsilon \) is denoted by \( \tau \) as illustrated in Fig. 2 (a). Since the desired EPI \( f \) to be reconstructed is densely sampled, it is apparent that \( \tau \geq d_{\text{range}} \) and \( d_{\text{range}} \) stands for the disparity range of the input decimated EPI \( \epsilon \), i.e., \( d_{\text{range}} = \max \epsilon - \min \epsilon \). It should be noted that a pre-shearing process relying on \( d_{\text{min}} \) is typically necessary for the input decimated EPI \( \epsilon \) in order to make sure that the new \( d'_{\text{min}} = 0 \) and \( d'_{\text{max}} = d_{\text{range}} \). Besides, the vertical sizes of the input decimated EPI \( \epsilon \) and reconstructed densely-sampled EPI \( f \) meet the condition that \( \bar{n} = (\bar{n} - 1)\tau + 1 \).

The reconstruction of the desired densely-sampled EPI \( f \) is typically performed via an iterative inpainting process with \( t \) iterations, corresponding to the intermediate reconstructed EPI result \( f_i, i \in [1, t] \cap \mathbb{Z} \). Besides, the shearlet analysis transform for reconstructing \( f \) is defined as \( S : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{q \times m \times n} \) and the shearlet synthesis transform is denoted by \( S^* : \mathbb{R}^{q \times m \times n} \rightarrow \mathbb{R}^{m \times n} \). Additionally, \( f_0 \) stands for the coarse estimation of \( f \), which is a zero-padded EPI for the input decimated EPI \( \epsilon \), i.e., \( f_0 : \tau, \cdot = \epsilon \) as shown in Fig. 2 (a). The reconstruction of \( f_i \) during iteration \( i \) is performed using the double relaxation method [4], which has been demonstrated to be faster and more robust than the original hard-thresholding algorithm in [3]:

\[
\begin{align*}
\hat{f}_i &= S^* \left( T_{\lambda_i} \left( S \left( f_i + \alpha \left( f_0 - M \circ f_i \right) \right) \right) \right),
\hat{f}_i &= \hat{f}_i + \beta_1 \left( \hat{f}_i - f_i - 1 \right), \\
\hat{f}_i &= \hat{f}_i + \beta_2 \left( \hat{f}_i - f_i - 2 \right),
\end{align*}
\]

where

\[
\begin{align*}
\beta_1 &= \frac{\text{sum} \left( \left( f_0 - \hat{f}_i \right) \circ M \circ \left( \hat{f}_i - f_i - 1 \right) \right)}{\text{sum} \left( \left( f_i - f_{i-1} \right) \circ M \circ \left( \hat{f}_i - f_i - 1 \right) \right)}, \\
\beta_2 &= \frac{\text{sum} \left( \left( f_0 - \hat{f}_i \right) \circ M \circ \left( \hat{f}_i - f_i - 2 \right) \right)}{\text{sum} \left( \left( f_i - f_{i-2} \right) \circ M \circ \left( \hat{f}_i - f_i - 2 \right) \right)}.
\end{align*}
\]

Here, \( \text{sum}(\cdot) \) returns the sum of all the elements in the input matrix, \( \alpha \) is a parameter for adjusting the convergence speed, \( \circ \) denotes the element-wise (Hadamard) product and \( M \) is a logical measuring matrix as shown in Fig. 2 (c), where ideally \( f_i \circ M = f_0 \). In addition, \( T_{\lambda_i}(\cdot) \) is a hard-thresholding operator [15] for the threshold value \( \lambda_i \), which linearly decreases from \( \lambda_{\text{max}} \) to \( \lambda_{\text{min}} \), with iteration \( i \) increasing from 1 to \( t \). As can be seen from (1) and (2), the computation time of the ST approach above is linearly dependent on the maximum iteration number \( t \). A reliable \( f_0 \), i.e., coarse estimation of \( f \), makes it feasible to accelerate ST with a smaller \( t \).

3.2. Mask-Accelerated Shearlet Transform (MAST)
In order to make a more reliable estimation of \( f_0 \) w.r.t. the desired densely-sampled EPI \( f \), one of the state-of-the-art learning-based optical flow methods, i.e., FlowNet2 [6], is utilized to estimate bidirectional flow between adjacent views in a horizontal-parallax SSLF, \( D_{\text{SSLF}} = \{ I_i | 1 \leq i \leq \bar{n} \} \), of which the corresponding unknown DSFL is denoted by \( D_{\text{dsfl}} = \{ I_r | 1 \leq r \leq \bar{n} \} \). Since a horizontal-parallax SSLF does not have vertical motions of image objects between any two neighboring views, only the horizontal component of the
optical flow displacement vector is kept after the bidirectional flow estimation. The bidirectional flow between \( \mathcal{I}_i \) and \( \mathcal{I}_{i+1} \) in the \( D^{\text{SSLF}} \) is represented by \( F_{i\rightarrow i+1} \) and \( F_{i+1\rightarrow i} \). A forward-backward consistency constraint [16] between \( F_{i\rightarrow i+1} \) and \( F_{i+1\rightarrow i} \) is applied here to roughly remove the inaccuracies caused by occlusions and large motions of image objects. Let \( \hat{r} = (r - 1)\%\tau \) and \( i = 1 + (r - \hat{r} - 1) / \tau \), the estimated bidirectional flow is then used to perform a coarse estimation of the parallax images in \( D^{\text{SSLF}} \) as follows:\(^2\):

\[
\hat{F}_{i\rightarrow j} = \begin{cases} 
\mathcal{I}_i & \text{for } \hat{r} = 0, \\
\mathcal{I}_i - \frac{\hat{r}}{\tau} F_{i\rightarrow i+1} & \text{for } 0 < \hat{r} < \frac{\tau}{2}, \\
\mathcal{I}_{i+1} - \frac{(r-\hat{r})}{\tau} F_{i+1\rightarrow i} & \text{for } \frac{\tau}{2} < \hat{r} < \tau, \\
0 & \text{for } \hat{r} = \frac{\tau}{2}.
\end{cases}
\]

Fig. 2. The coarse estimation, measuring matrix (soft mask) and estimation refinement for a densely-sampled EPI \( f \).

Here, \( g(\cdot, \cdot) \) is an inverse warping function using bicubic interpolation [17]. The roughly-estimated \( D^{\text{SSLF}} \) is then turned into densely-sampled EPIs, such that the coarse estimation \( \hat{f}_0 \) of \( f \) is partially restored as displayed in Fig. 2(b). Note that the large missing areas are caused by the filtering of the unreliable optical flows using the bidirectional consistency check. However, the roughly inpainted areas in \( f_0 \) are not accurate enough for directly using ST. Specifically, due to the accumulation error of the optical flow in the interpolation algorithm in (3), horizontal lines of \( f_0 \) near the locations, i.e., \( \hat{r} = \frac{\tau}{2} \), have larger inpainting errors than those near the ground-truth regions, i.e., \( \hat{r} = 0 \). Therefore, a novel ST-based method, Mask-Accelerated Shearlet Transform (MAST), is proposed to solve this problem by replacing the measuring matrix in (1) and (2) with an elaborately-designed soft mask, i.e.

\[
M(r, c) = \begin{cases} 
1.0 & \text{for } \hat{r} = 0, \\
\omega (1 - \frac{2\hat{r}}{\tau})^2 & \text{for } \hat{r} > 0, \quad f_0(r, c) > 0, \\
0 & \text{for } \hat{r} > 0, \quad f_0(r, c) = 0,
\end{cases}
\]

where \( \omega \in (0, 1), r \in [1, \hat{n}] \cap \mathbb{Z} \) and \( c \in [1, m] \cap \mathbb{Z} \). An example soft mask corresponding to \( f_0 \) is illustrated in Fig. 2 (d). It can be seen that this mask suppresses the contributions of \( f_0 \) in the regions which are not inpainted or meet the condition that \( \hat{r} \) is close to \( \frac{\tau}{2} \); however, it enhances the contributions from the ground-truth nearby areas, thus effectively improving the initialization of the densely-sampled EPIs for the iterative double relaxation-based ST in Sect. 3.1.

4. EXPERIMENTS

4.1. Experimental settings

Datasets. The high density camera array dataset [18] is a real-world 4D light field dataset that can be utilized to evaluate light field angular super-resolution methods with large disparity ranges. Nine different scenes in this dataset are captured by a high-resolution and high-definition DSLR camera in a precise gantry system, such that nine corresponding light field sub-datasets are built. Eight of these sub-datasets have an angular resolution of \( 101 \times 21 \). The remaining one has an angular resolution of \( 99 \times 21 \). The spatial resolution of all the sub-datasets is \( 3976 \times 2652 \) pixels. The raw images in each sub-dataset have black areas near the image borders, which is due to the calibration, and large disparities between neighboring views, which make it difficult to use these raw images as ground-truth light field data directly. To overcome this limitation, a cutting and scaling strategy is proposed as shown in Fig. 3 (j). In particular, a bottom-right 16:9 image is cut from a raw image with preserving 95% of the width of this raw image. The cut image is then resized to \( 1024 \times 576 \) pixels using bicubic interpolation. Finally, the top 97 images after the process of the cutting and scaling strategy for each sub-dataset are kept and used as the ground-truth horizontal-parallax light field dataset \( D_\mu, \mu \in [1, 9] \cap \mathbb{Z} \). In other words, \( D_\mu = \{ I^\mu_j \mid 1 \leq j \leq n \}, I^\mu_j \in \mathbb{R}^{m \times n} \), where \( n = 97, m = 1024 \) and \( N = 576 \). The middle image, i.e., \( I^9_9 \), of each ground-truth 3D light field dataset \( D_\mu \) is exhibited in Fig. 3 (a)-(i). The SSLF \( D_\mu^{\text{SSLF}} \) from \( D_\mu \) is generated by using an interpolation rate \( \delta (= 16) \) as shown in Fig. 1 (a) and (c), such that \( D_\mu^{\text{SSLF}} = \{ I^\mu_i \mid 1 \leq i \leq \hat{n} \}, \hat{n} = (n-1)/\delta + 1 \).

\(^1\)Here, ‘\%’ stands for the modulo operation.

\(^2\)Assume that \( \tau \% 2 = 0 \) for this \( D^{\text{SSLF}} \).
The minimum and average per-view PSNR results (in dB, explained in Sect. 4.1) for the performance evaluation of different DSLF reconstruction methods on nine light field evaluation datasets.

<table>
<thead>
<tr>
<th>Method</th>
<th>$\mathcal{D}_1$</th>
<th>$\mathcal{D}_2$</th>
<th>$\mathcal{D}_3$</th>
<th>$\mathcal{D}_4$</th>
<th>$\mathcal{D}_5$</th>
<th>$\mathcal{D}_6$</th>
<th>$\mathcal{D}_7$</th>
<th>$\mathcal{D}_8$</th>
<th>$\mathcal{D}_9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAST</td>
<td>30.167</td>
<td>22.965</td>
<td>26.866</td>
<td>27.920</td>
<td>29.541</td>
<td>32.448</td>
<td>23.119</td>
<td>23.847</td>
<td>29.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>$\mathcal{D}_1$</th>
<th>$\mathcal{D}_2$</th>
<th>$\mathcal{D}_3$</th>
<th>$\mathcal{D}_4$</th>
<th>$\mathcal{D}_5$</th>
<th>$\mathcal{D}_6$</th>
<th>$\mathcal{D}_7$</th>
<th>$\mathcal{D}_8$</th>
<th>$\mathcal{D}_9$</th>
</tr>
</thead>
</table>

Table II. The average computation time of reconstructing a densely-sampled EPI (RGB channels) using ST and MAST.

<table>
<thead>
<tr>
<th>Method</th>
<th>$\tau = 32$</th>
<th>$\tau = 48$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAST</td>
<td>2.813</td>
<td>5.073</td>
</tr>
</tbody>
</table>

The DSLF to be reconstructed is $\mathcal{D}_{\mu}^{\text{dsL}} = \{ \tilde{T}_\mu^r | 1 \leq r \leq \bar{n} \}$, $\bar{n} = (\bar{n} - 1) \tau + 1$ as described in Sect. 3.

Disparity estimation. The horizontal disparities between neighboring views in each $\mathcal{D}_{\mu}^{\text{ssL}}$ are calculated via the optical flow algorithm in Sect. 3.2. The estimated minimum disparity $d_{\min}$, maximum disparity $d_{\max}$ and disparity range $d_{\text{range}}$ are illustrated in Fig. 5. The sampling interval $\tau$ should be as small as possible in order to save computation time for both ST and MAST, while it has two constraints that $\tau \% 6 = 0$ and $\tau \geq d_{\text{range}}$ (see Sect. 3.1). Therefore, it can be seen from the figure that the best sampling interval $\tau$ for datasets $\mathcal{D}_{\mu}$, $\mu \in \{1, 2, 7\}$ is 32 and for the other six datasets, $\tau = 48$.

Evaluation criteria. The per-view PSNR for a ground-truth dataset $\mathcal{D}_\mu$ and the reconstructed $\mathcal{D}_{\mu}^{\text{dsL}}$ from $\mathcal{D}_{\mu}^{\text{ssL}}$ for it is described as below:

$$
\text{MSE}_{j}^{\mu} = \frac{1}{3 \cdot m \cdot N} \sum_{x=1}^{m} \sum_{y=1}^{N} \left( \frac{1}{\bar{n}} \sum_{r=1}^{\bar{n}} T_{\mu, \omega(r)}(x, y) - T_{\mu}(x, y) \right)^2,
$$

$$
\text{PSNR}_{j}^{\mu} = 10 \log_{10} \left( \frac{255^2}{\text{MSE}_{j}^{\mu}} \right).
$$

(5)

The minimum and average per-view PSNRs constitute the evaluation criteria for the evaluation of different DSLF reconstruction methods on a dataset $\mathcal{D}_{\mu}$.

Implementation details. For a dataset $\mathcal{D}_{\mu}$, the construction of a specifically-designed universal shearlet system [3] with $\xi$ scales relies on the sampling interval $\tau$ of it, i.e., $\xi = \lceil \log_2 \tau \rceil$. The parameter $\omega$ in (4) is set to 0.1. The maximum threshold value $\lambda_{\text{max}}$ and minimum threshold value $\lambda_{\text{min}}$ are set to 8 and 0.04, respectively. Note that these two values are for the case of using a normalized coarsely-sampled EPI $\varepsilon$, i.e.,

$$
\varepsilon = \varepsilon - \min(\varepsilon) \quad \text{max}(\varepsilon) - \min(\varepsilon),
$$

(6)

where $\max(\cdot)$ and $\min(\cdot)$ return the maximum value and the minimum value of an input matrix, respectively. The reconstructed $f$ using this normalized $\varepsilon$ is then rescaled back to original range of values via

$$
f = (\max(\varepsilon) - \min(\varepsilon)) f + \min(\varepsilon).
$$

(7)

Besides, for the maximum iteration number of ST, $t_{\text{max}} = 100$ and for that of MAST, $t_{\text{max}} = 50$. Regarding the parameter controlling the convergence speed in (1), $\alpha = 0.9$. Both ST and MAST are implemented by using CUDA and executed on an Nvidia GeForce GTX Titan X 12 GB GPU.

4.2. Results and analysis

The proposed method and baseline approaches are evaluated quantitatively and qualitatively as follows:

Quantitative evaluation. The minimum and average per-view PSNR values of using different DSLF reconstruction methods on different horizontal-parallax light field datasets...
are presented in Table I. It can be seen from the minimum per-view PSNR data that the proposed MAST method achieves the best performance on most of the datasets except for D_6 and D_7. However, on these two datasets, the minimum per-view PSNR values of ST are only 0.037 dB and 0.067 dB higher than those of MAST. With regard to the average per-view PSNR data at the bottom of Table I, MAST still outperforms the other baseline methods on most datasets except for D_4, D_7 and D_8, which demonstrates the effectiveness of the proposed DSLF reconstruction approach again. The computation efficiency of both ST and MAST is evaluated in terms of computation time as shown in Table II. The proposed MAST is significantly faster than ST, i.e. MAST requires only ≈ 35% computation time of ST, which is mainly because MAST has 30 refinement iterations while ST needs 100 iterations for the DSLF reconstruction on the challenging horizontal-parallax real-world light field datasets. This also demonstrates that MAST is much more efficient than ST for DSLF reconstruction. Note that the computation time of optical flow estimation and inverse warping parts of MAST can be ignored compared with the computation time of the iterative EPI-refinement process since both of them are performed in real-time.

**Qualitative evaluation.** The minimum per-view PSNR data for all the DSLF reconstruction methods on different light field evaluation datasets are plotted in Fig. 6. It is apparent that both SepConv and PIASC have almost the same minimum per-view PSNR results on all the datasets, which are much lower than ST and MAST. This suggests that the two DSLF reconstruction methods using the state-of-the-art video frame interpolation technology are not appropriate for DSLF reconstruction from SSLFs with large disparity ranges. Besides, the proposed MAST approach outperforms ST on most of the challenging light field datasets, which indicates that MAST is more effective than ST for DSLF reconstruction. The reconstructed images using different DSLF reconstruc-
tion methods are visualized and compared in Fig. 4. Since SepConv and PIASC have similar DSLF reconstruction performance, only SepConv is compared here. For the top row, the image parts of the checkerboard and Siemens star on $T_{23}^2$ of $D_1$ are chosen as the interesting areas to be compared. It can be seen from Fig. 4 (b) that the reconstructed checkerboard using SepConv has serious blur artifacts, which is mainly because the the size of repetitive check patterns of the checkerboard is much smaller than the disparities of it, such that SepConv is incapable of knowing the true motion of this checkerboard. As can be seen from Fig. 4 (c), the recovered checkerboard using ST is slightly better than that of using SepConv, while the reconstructed Siemens star has obvious artifacts. In Fig. 4 (d), the proposed MAST method achieves the best reconstruction performance with visually correct and sharp results, which proves the effectiveness of the proposed MAST method composed of optical-flow-based coarse estimation and mask-assisted iterative estimation refinement for EPIs. Regarding the bottom row Fig. 4, part of the tablecloth with foreground and the Fraunhofer IIS logo are selected as the interesting areas from $T_{25}^6$ of $D_6$. Both of the reconstructed results in Fig. 4 (f) using SepConv are blur, which, on the one hand, is caused by the small size of the repetitive pattern of the tablecloth; on the other hand, the size of the convolution kernels of SepConv is only $51 \times 51$, restricting the performance of it in handling DSLF reconstruction from SSLFs with large disparity ranges. The DSLF reconstruction results of ST in Fig. 4 (g) do not have this kind of “blur” problem. However, the reconstructed tablecloth area has evident artifacts, which are well handled by the proposed MAST method as illustrated in Fig. 4 (h). It implies that the optical-flow-based coarse estimation and mask-assisted iterative estimation refinement in MAST are beneficial to improving the final DSLF reconstruction performance.

5. CONCLUSION

This paper presents a novel coarse-to-fine method, MAST, for DSLF reconstruction from SSLFs with large disparity ranges. The proposed MAST method fully leverages a state-of-the-art optical flow estimation method, i.e., FlowNet2, to roughly estimate a densely-sampled EPI from a sparsely-sampled EPI. Based on the coarsely-inpainted densely-sampled EPI and the inevitable error accumulation of any optical flow algorithm, a soft mask is elaborately designed for the iterative hard-thresholding-based estimation refinement approach in ST. Experimental results show that MAST achieves better performance than the other state-of-the-art DSLF reconstruction methods on nine challenging real-world horizontal-parallax light field datasets with large disparity ranges (up to 35 pixels). Moreover, MAST is a time-efficient algorithm that is nearly three times faster than ST.

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. We thank Nvidia for their GPU donation.

6. REFERENCES