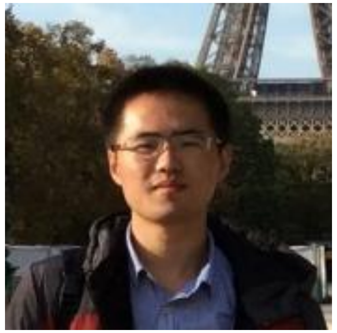
FAST: Flow-Assisted Shearlet Transform for Densely-sampled Light Field Reconstruction

Yuan Gao, Reinhard Koch, Robert Bregovic and Atanas Gotchev

CAU
Christian-Albrechts-Universität zu Kiel

Tampereen yliopisto
Tampere University

Outline

- Introduction
- Related Work
  - Shearlet Transform (ST)
  - Optical Flow
- Methodology
  - Motivation
  - Flow-Assisted Shearlet Transform (FAST)
- Experiments
- Conclusion

Introduction

- Densely-Sampled Light Field (DSLF)
  - A 4D approximation of the plenoptic function [1]
  - Disparities between adjacent views are less than one pixel [2]

[1] M. Levoy and P. Hanrahan, “Light field rendering,” in SIGGRAPH, 1996, pp. 31–42.
[2] S. Vagharshakyan, R. Bregovic, and A. Gotchev, “Image based rendering technique via sparse representation in shearlet domain,” in ICIP, 2015, pp. 1379–1383.

Introduction

- DSLF-based Research [3]
  - Depth estimation
  - Super-resolution
  - Synthetic aperture imaging

  …

[3] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu, “Light field image processing: An overview,” IEEE J-STSP, vol. 11, no. 7, pp. 926–954, 2017.

Introduction

- DSLF's applications
  - 3D-TV [4]
  - Virtual Reality (VR) [5]

[4] A. Smolic, “3D video and free viewpoint video - from capture to display,” Pattern Recognition, vol. 44, no. 9, pp. 1958–1968, 2011.
[5] J. Yu, “A light-field journey to virtual reality,” IEEE MultiMedia, vol. 24, no. 2, pp. 104–112, 2017.

Introduction

- DSLF reconstruction from Sparsely-Sampled Light Fields (SSLFs)

Interpolation rate

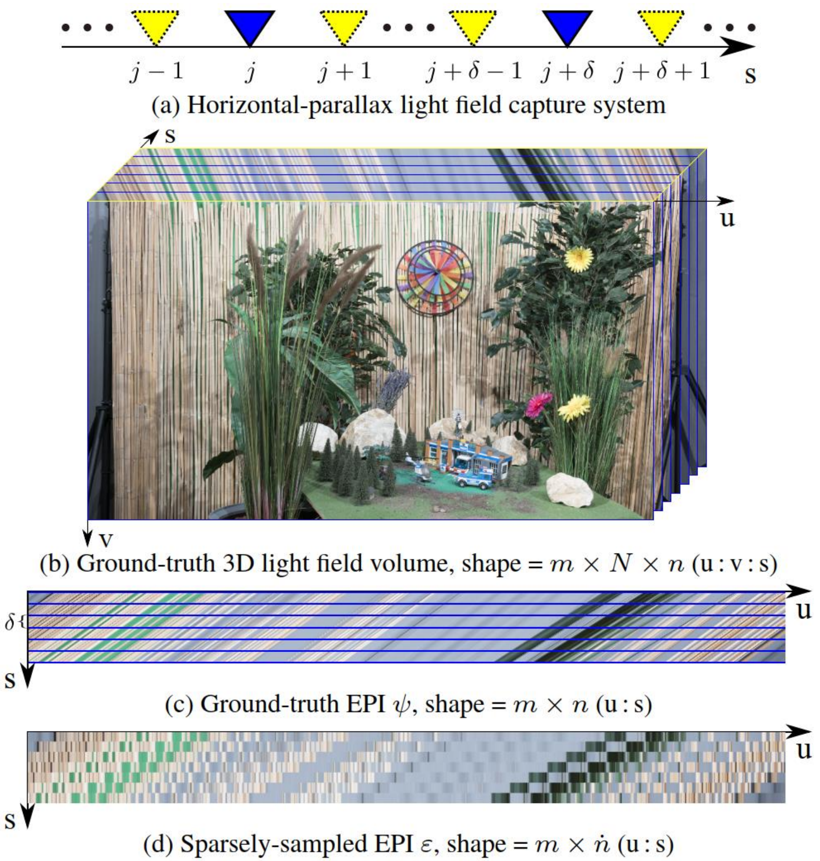


Fig. 1. Introduction to the DSLF reconstruction problem.

Outline

- Introduction
- Related Work
  - Shearlet Transform (ST)
  - Optical Flow
- Methodology
  - Motivation
  - Flow-Assisted Shearlet Transform (FAST)
- Experiments
- Conclusion

Related Work (ST)

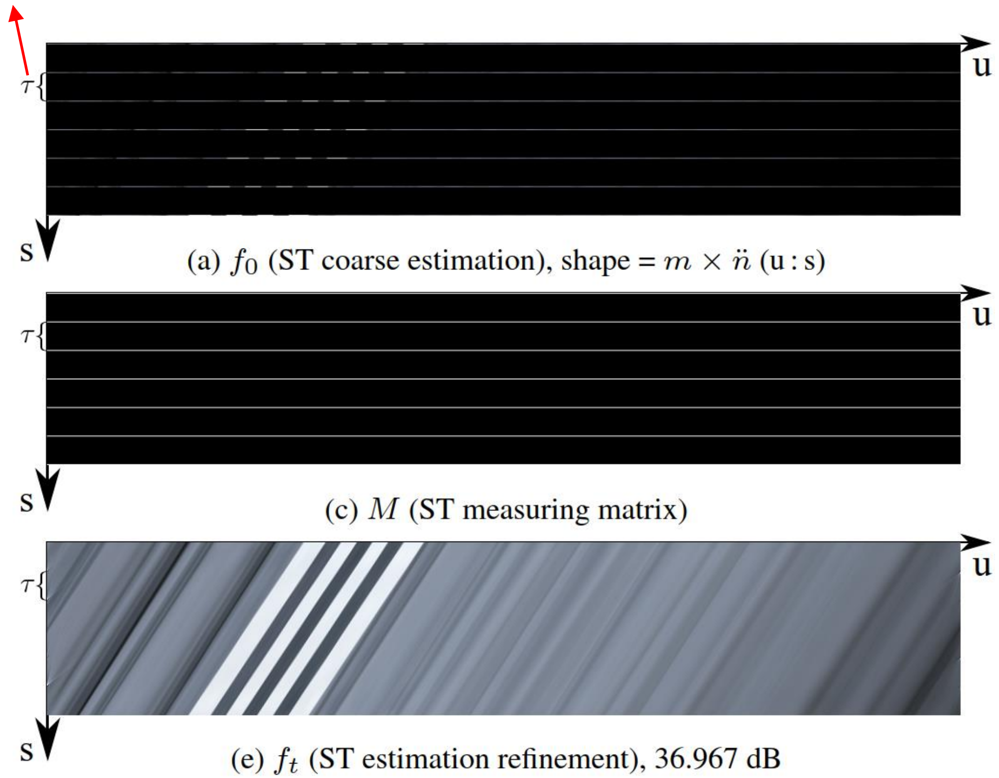
- Shearlet Transform (ST) [6, 7]
  - Pre-shearing
  - Shearlet system construction
  - Sparsity regularization
  - Post-shearing

[6] S. Vagharshakyan, R. Bregovic, and A. Gotchev, "Light field reconstruction using shearlet transform," IEEE TPAMI, vol. 40, no. 1, pp. 133–147, 2018.
[7] S. Vagharshakyan, R. Bregovic, and A. Gotchev, "Accelerated shearlet-domain light field reconstruction," IEEE J-STSP, vol. 11, no. 7, pp. 1082–1091, 2017.

Related Work (ST)

- Sparsity regularization
  - Analysis transform S
  - Hard thresholding T_{λ_i}
  - Synthesis transform S*
  - Double overrelaxation (DORE, optional)

$$\hat{f}_i = S^*\left(T_{\lambda_i}\left(S\left(f_i + \alpha(f_0 - M \circ f_i)\right)\right)\right)$$



(a) f_0 (ST coarse estimation), shape = $m \times \ddot{n}$ (u : s)

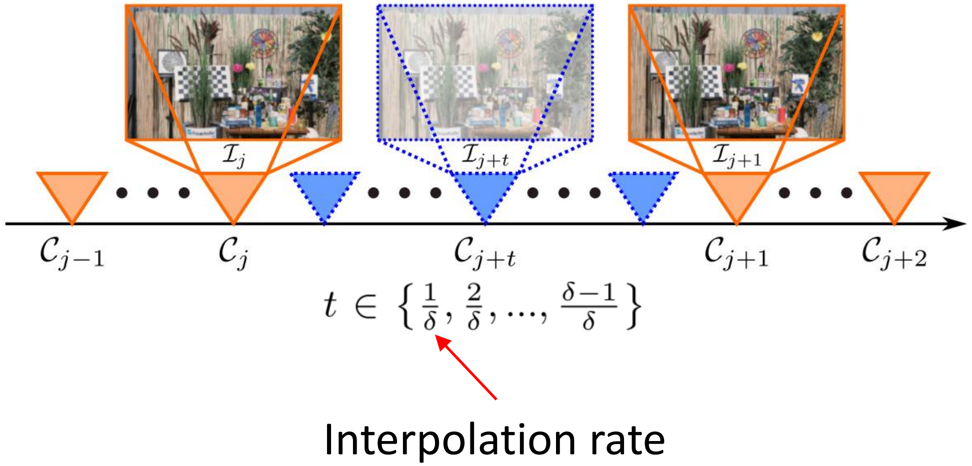
(c) M (ST measuring matrix)

(e) f_t (ST estimation refinement), 36.967 dB

Related Work (Optical Flow)

- View synthesis using optical flow
  - Backward warping
  - Image blending with a soft visibility map ($\mathcal{V}_{t\leftarrow 0}$)

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



- Challenge
  - It is difficult to calculate inverse optical flows, $\mathcal{F}_{t\to 0}$ and $\mathcal{F}_{t\to 1}$
  - They are approximated from bidirectional optical flows, $\mathcal{F}_{0\to 1}$ and $\mathcal{F}_{1\to 0}$

$$\tilde{\mathcal{F}}_{t\to 0} = -(1-t)t\mathcal{F}_{0\to 1} + t^2\mathcal{F}_{1\to 0},$$
$$\tilde{\mathcal{F}}_{t\to 1} = (1-t)^2\mathcal{F}_{0\to 1} - t(1-t)\mathcal{F}_{1\to 0}. \qquad (2)$$

[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.

Outline

- Introduction
- Related Work
  - Shearlet Transform (ST)
  - Optical Flow
- Methodology
  - Motivation
  - Flow-Assisted Shearlet Transform (FAST)
- Experiments
- Conclusion

Methodology

- Motivation
  - ST is especially effective in reconstructing a DSLF from a SSLF with a large disparity range (> 16 pixels) [9]
  - ST requires a precise disparity estimation of the input SSLF, which can be solved by using a state-of-the-art optical flow method
  - The qualities of estimated bidirectional disparity maps are not good enough to perform DSLF reconstruction
  - How to combine ST-based and disparity-based DSLF reconstruction methods?

[9] Y. Gao, R. Koch, R. Bregovic, and A. Gotchev, “IEST: Interpolation-enhanced shearlet transform for light field reconstruction using adaptive separable convolution,” in EUSIPCO, 2019.

Methodology

- Flow-Assisted Shearlet Transform (FAST)
  - Disparity Refinement Network (DRN)
    - 6 hierarchies in the encoder part and 5 hierarchies in the decoder part
  - View Synthesis Network (VSN)
    - 4 hierarchies in the encoder part and 3 hierarchies in the decoder part

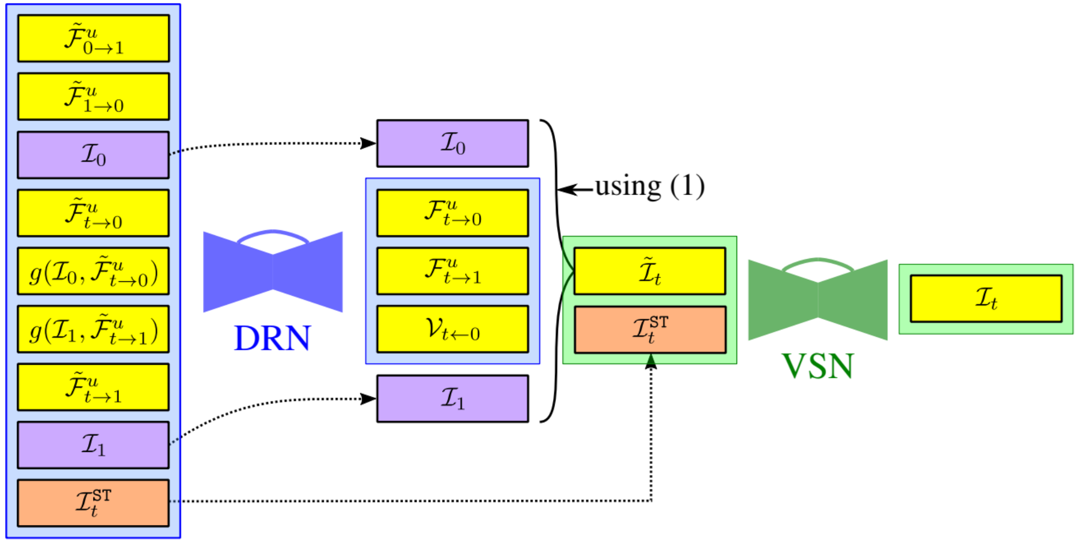


Fig. 2. An overview of the architecture of FAST. Note that $\mathcal{I}_t^{\mathrm{ST}}$ is the view reconstructed by ST and $t \in \left\{\frac{1}{\delta}, \frac{2}{\delta}, \ldots, \frac{\delta-1}{\delta}\right\}$ (see Sect. 1).

Methodology

- Flow-Assisted Shearlet Transform (FAST)
  - Loss functions
    - VSN reconstruction loss, $\mathcal{L}^{\mathrm{VSN}}$
    - DRN reconstruction loss, $\mathcal{L}^{\mathrm{DRN}}$
    - Warping loss, $\mathcal{L}^{\mathrm{W}}$

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

where

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

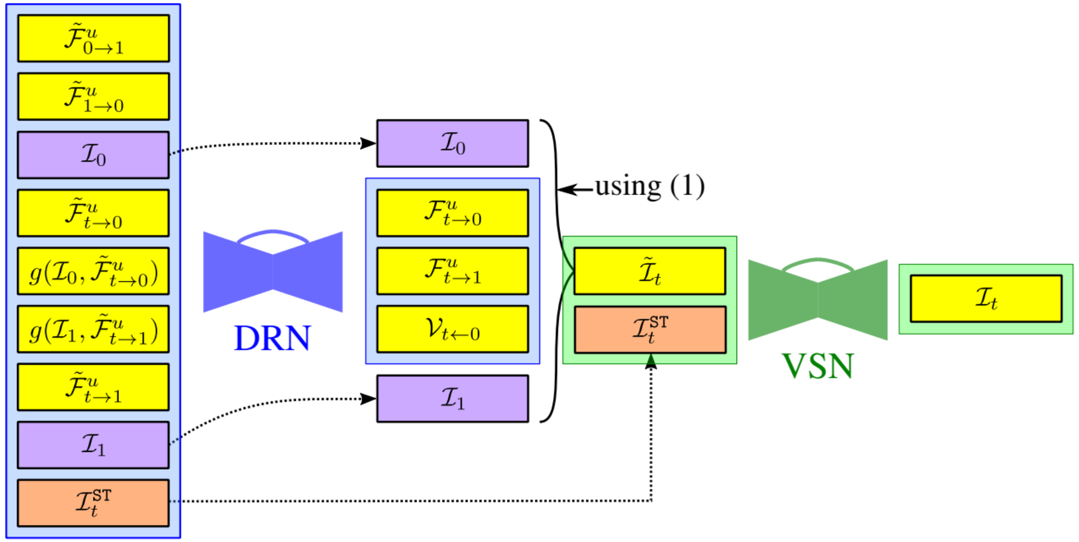


Fig. 2. An overview of the architecture of FAST. Note that $\mathcal{I}_t^{\mathrm{ST}}$ is the view reconstructed by ST and $t \in \left\{\frac{1}{\delta}, \frac{2}{\delta}, \ldots, \frac{\delta-1}{\delta}\right\}$ (see Sect. 1).

Outline

- Introduction
- Related Work
  - Shearlet Transform (ST)
  - Optical Flow
- Methodology
  - Motivation
  - Flow-Assisted Shearlet Transform (FAST)
- Experiments
- Conclusion

Experiments

- Training datasets
  - Standford light field dataset captured by a Lego gantry
  - 4D light field benchmark with synthetic light fields [10]
  - 220 horizontal-parallax sub-datasets with same spatial resolution of 512 × 512 pixels

[10] 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.

Experiments

- Evaluation dataset
  - High Density Camera Array (HDCA) [11]

(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.

[11] 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.

Experiments

- Evaluation dataset
  - High Density Camera Array (HDCA) [11]
  - D_{μ}: Ground-truth horizontal-parallax light fields (n × 512 × 320 × 3 pixels)
    - $n = 97$
  - Interpolation rate $\delta = 32$
  - S_{μ}: Input SSLFs ($\dot{n}$ × 512 × 320 × 3 pixels)
    - $\dot{n} = \left(\frac{n-1}{\delta} + 1\right) = 4$

[11] 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.

Experiments

- Disparity estimation (for evaluation dataset)
  - Interpolation rate δ = 32
  - Use a state-of-the-art optical flow method, PWC-Net [12]
  - Densely-sampled, $\frac{d_{range}}{\delta} < 1$ pixel

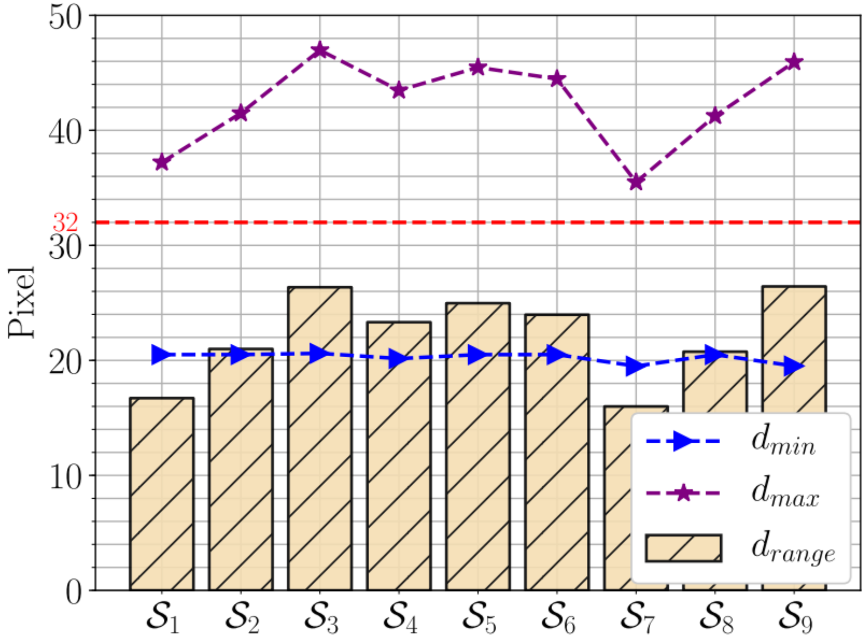


Fig. 4. Disparity estimations of $\mathcal{S}_\mu$ $(1 \leqslant \mu \leqslant 9)$ using PWC-Net.

[12] 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.

Experiments

- Evaluation criteria
  - Minimum per-view PSNR

$$\mathrm{MSE}_{i,\mu} = \frac{1}{3 \times M \times N} \sum_{x=1}^{M} \sum_{y=1}^{N} \left\| \tilde{\mathcal{I}}^{d}_{i,\mu}(x,y) - \mathcal{I}^{d}_{i,\mu}(x,y) \right\|_2^2,$$

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

Experiments

- Implementation details
  - ST
    - Iteration number = 100
    - Parameter α = 20
    - Use a low-pass initial estimation
    - Use the DORE algorithm
  - Use an Nvidia Titan Xp GPU
  - Min-batch size: 6
  - Learning rate: 0.0001
  - Epoch number: 1,500
  - Training time: 36 hours

$$\hat{f}_i = S^*\left(T_{\lambda_i}\left(S\left(f_i + \alpha(f_0 - M \circ f_i)\right)\right)\right)$$

Experiments

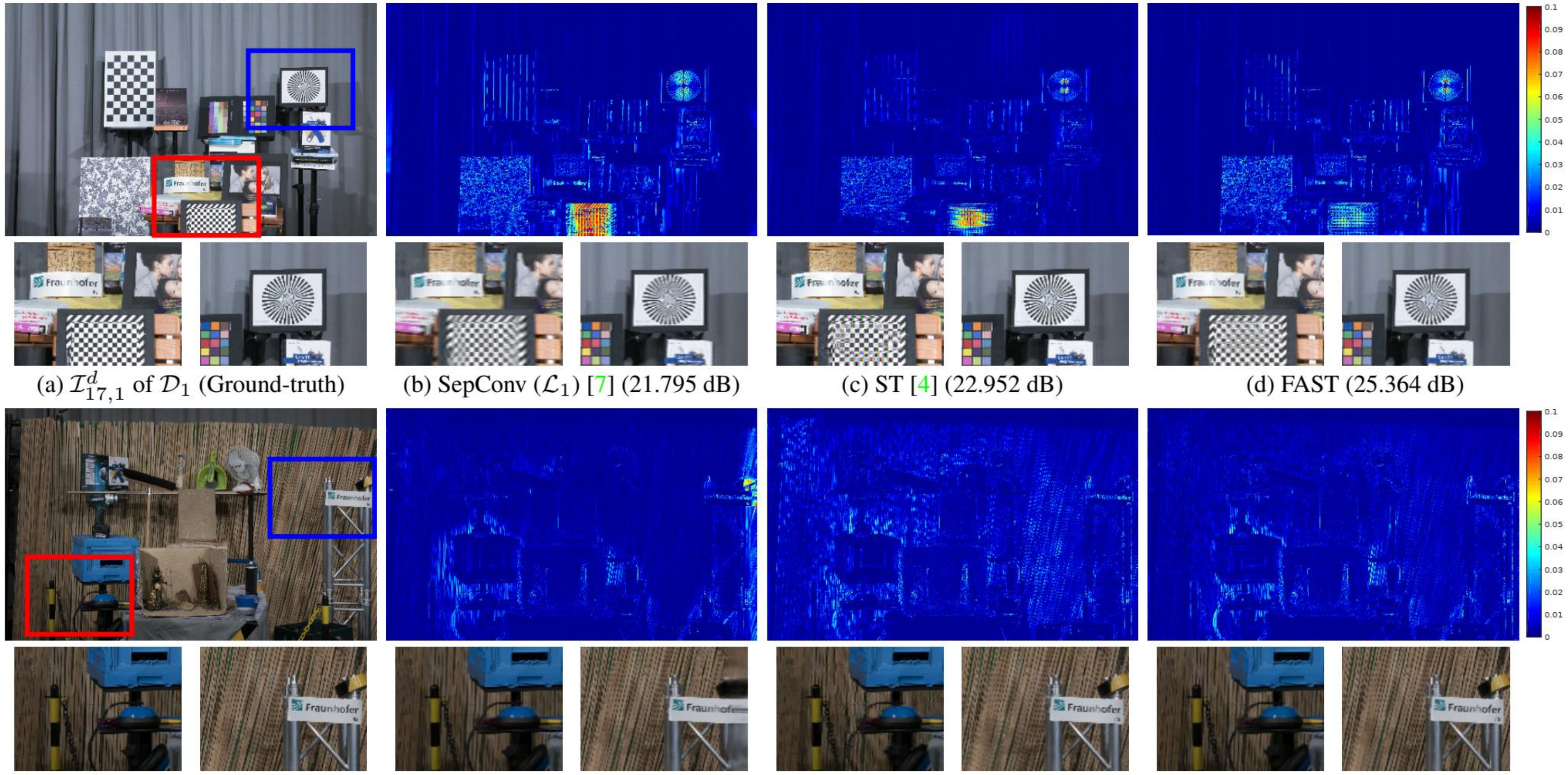
- Quantitative evaluation

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**

Experiments

- Qualitative evaluation



(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.

Outline

- Introduction
- Related Work
  - Shearlet Transform (ST)
  - Optical Flow
- Methodology
  - Motivation
  - Flow-Assisted Shearlet Transform (FAST)
- Experiments
- Conclusion

Conclusion

- FAST performs DSLF reconstruction by means of
  - PWC-Net, a state-of-the-art optical flow method
  - ST, a state-of-the-art DSLF reconstruction method
  - Super-SloMo, a state-of-the-art video frame interpolation method
- FAST outperforms the other state of the arts on nine challenging horizontal-parallax real-world DSLF sub-datasets for varying large disparity ranges (16 - 26 pixels)