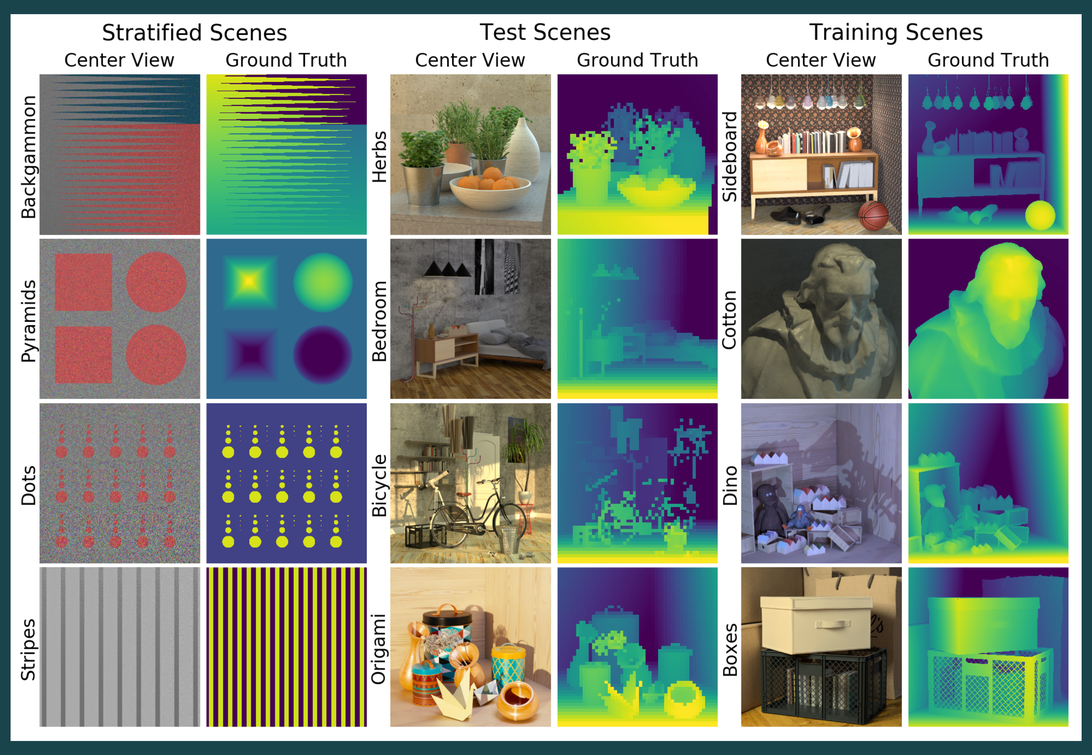
A Benchmark for Depth Estimation on 4D Light Fields

Katrin Honauer[1], Ole Johannsen[2], Daniel Kondermann[1], and Bastian Goldluecke[2]

[1]HCI, Heidelberg University, {firstname.lastname}@iwr.uni-heidelberg.de
[2]University of Konstanz, {firstname.lastname}@uni-konstanz.de
The first two authors contributed equally.

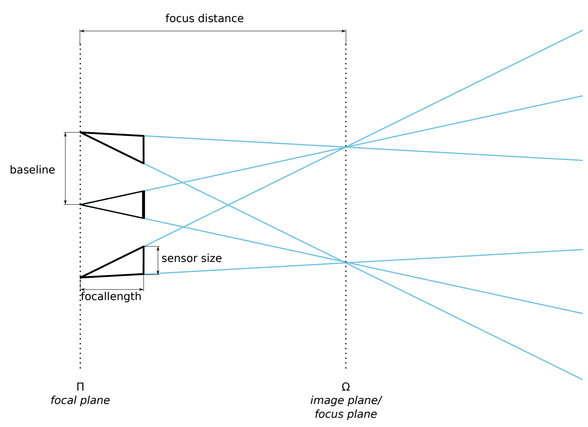
Contribution

- A dataset and evaluation methodology for depth estimation on 4D light fields
- **Four stratified scenes**: each evaluates an individual challenge and corresponding metric,
  - Backgammon: occlusions and fine structures
  - Pyramids: bumpiness on continuous planar and non-planar surfaces
  - Dots: fine structures and precision under the influence of noise
  - Stripes: sensitivity to low texture and contrast at occlusions
- **20 photorealistic scenes**: region specific evaluation at discontinuities, fine structures, and smooth surfaces
- Website: **www.lightfield-analysis.net**
  - Download scenes, evaluation toolkit, and Blender addon to render lightfields
  - Analyse algorithms on the benchmark table and with interactive visualizations



Dataset

- Available data per scene: 8bit light fields (9x9x512x512x3), camera parameters, 16bit ground truth disparities (4 hidden test scenes)
- Dataset creation with Blender: internal renderer for the stratified scenes, Cycles renderer for the photorealistic scenes
- Regular grid of views, identical cameras with parallel optical axes
- Cameras and image planes are shifted such that all cameras capture the same area at a certain focus distance. The resulting disparity is within around $-2px$ to $2px$.



Benchmark

Algorithms

We compare 4 state-of-the-art algorithms and one unpublished generic multiview stereo lab code implementation to test the benchmark and obtain an initial evaluation. The algorithms are LF [1], LF_OCC [3], EPI1 [2], and EPI2 [4].

Metrics

MSE: Mean Squared Error *100

$$\text{MSE}_{\mathcal{M}} = \frac{\sum_{x \in \mathcal{M}} (d(x) - gt(x))^2}{|\mathcal{M}|} * 100$$

Bumpiness: Smoothness on continous surfaces

$$\text{Bumpiness} = \frac{\sum_{x \in \mathcal{M}} \min(0.05, \|\mathrm{H}_f(x)\|_F)}{|\mathcal{M}|} * 100,$$

where $f = d - gt$ and H is the Hessian matrix.

BadPix: Ratio of pixels in area $\mathcal{M}$ with an error above a threshold t

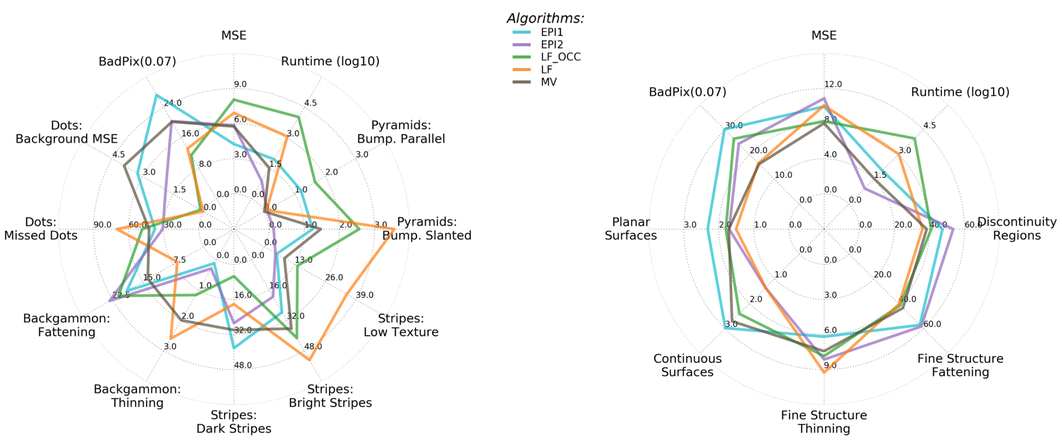
$$\text{BadPix}_{\mathcal{M}}(t) = \frac{|\{x \in \mathcal{M} \colon |d(x) - gt(x)| > t\}|}{|\mathcal{M}|}$$

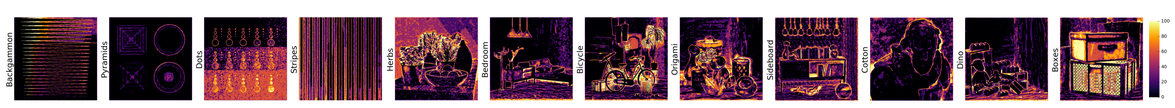
Fattening: Pixels at an occlusion boundary that are wrongly classified as foreground

$$\text{Fattening}_{\mathcal{M}}(t) = \frac{|\{x \in \mathcal{M} \colon gt(x) - d(x) < t\}|}{|\mathcal{M}|}$$

Thinning is defined analogously.

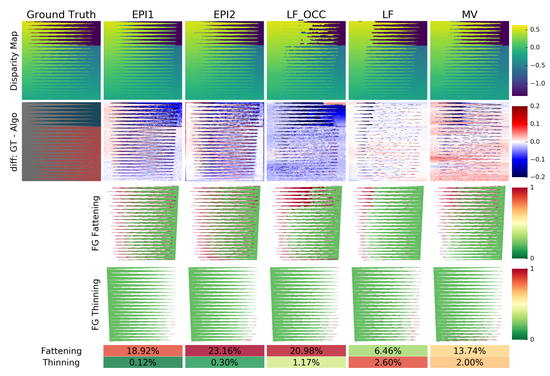
Visualizations



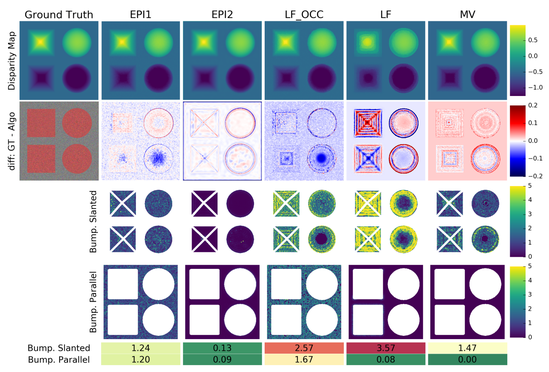


Evaluation on Stratified Scenes

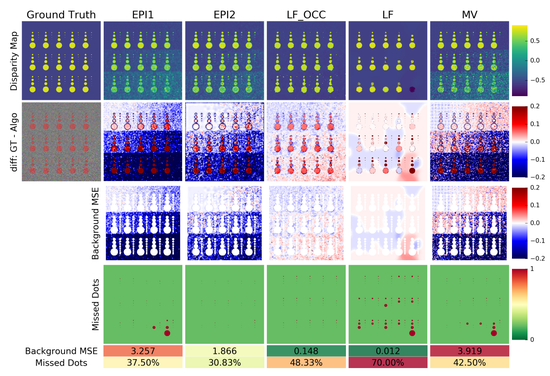
Backgammon: occlusions and fine structure



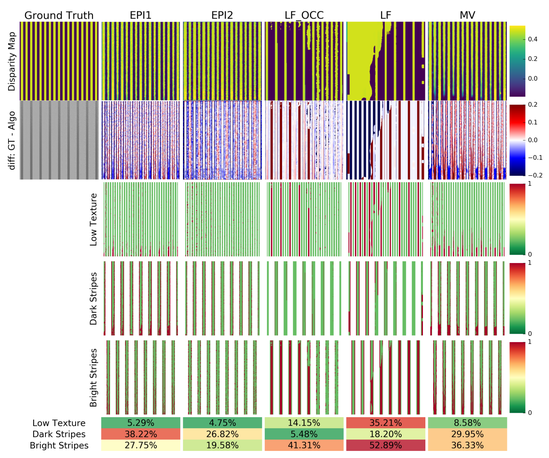
Pyramids: bumpiness on continuous surfaces



Dots: noise

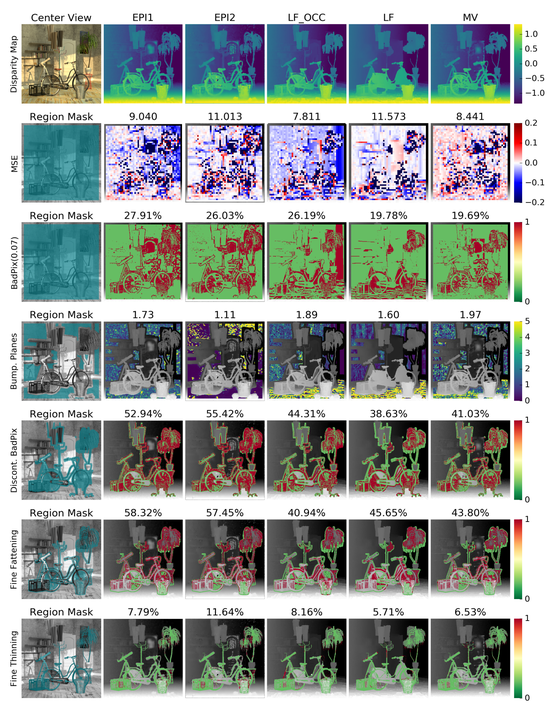


Stripes: low texture

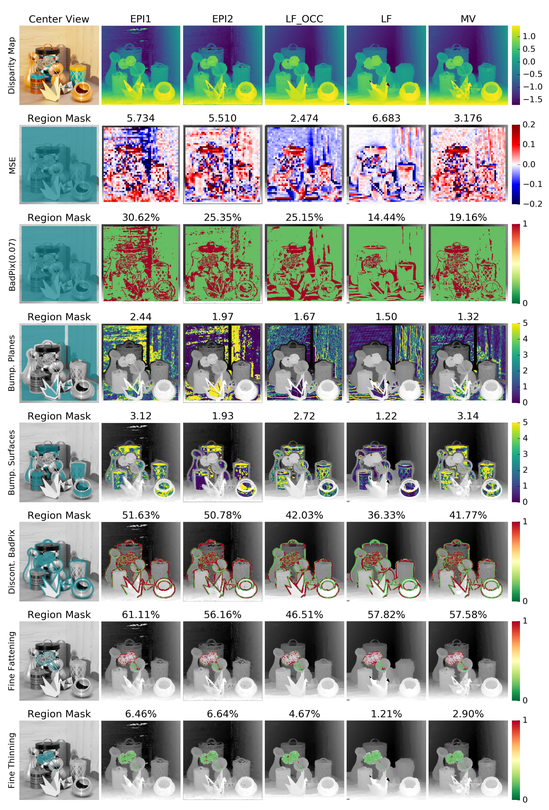


Evaluation on Photorealistic Scenes

Bicycle



Origami



References

[1] H. Jeon, J. Park, G. Choe, J Park, Y. Bok, Y. Tai, and I. Kweon. Accurate depth map estimation from a lenslet light field camera. In *Proc. CVPR*, 2015.
[2] Ole Johannsen, Antonin Sulc, and Bastian Goldluecke. What sparse light field coding reveals about scene structure. In *Proc. CVPR*, 2016.
[3] T. Wang, A. Efros, and R. Ramamoorthi. Occlusion-aware depth estimation using light-field cameras. In *Proc. ICCV*, 2015.
[4] S. Wanner and B. Goldluecke. Reconstructing reflective and transparent surfaces from epipolar plane images. In *Proc. GCPR*, 2013.

This work was supported by the ERC Starting Grant "Light Field Imaging and Analysis" (LIA 336978, FP7-2014).
Presented at the Asian Conference on Computer Vision (ACCV), Taipei, November 2016.