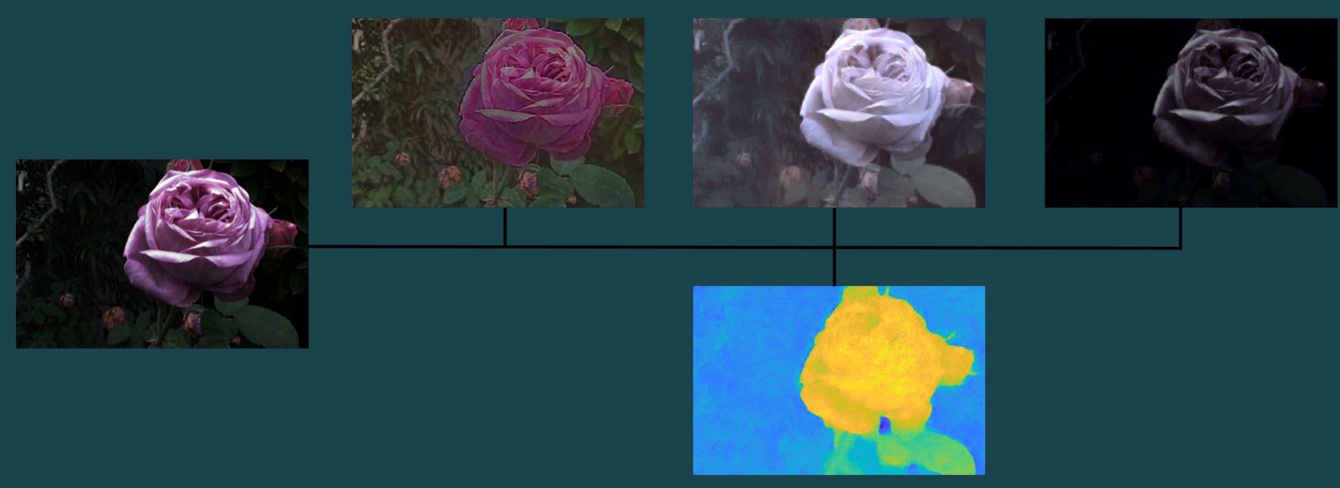
Intrinsic Light Field Decomposition and Disparity Estimation with Deep Encoder-Decoder Network

Anna Alperovich, Ole Johannsen and Bastian Goldluecke
University of Konstanz
https://www.cvia.uni-konstanz.de/

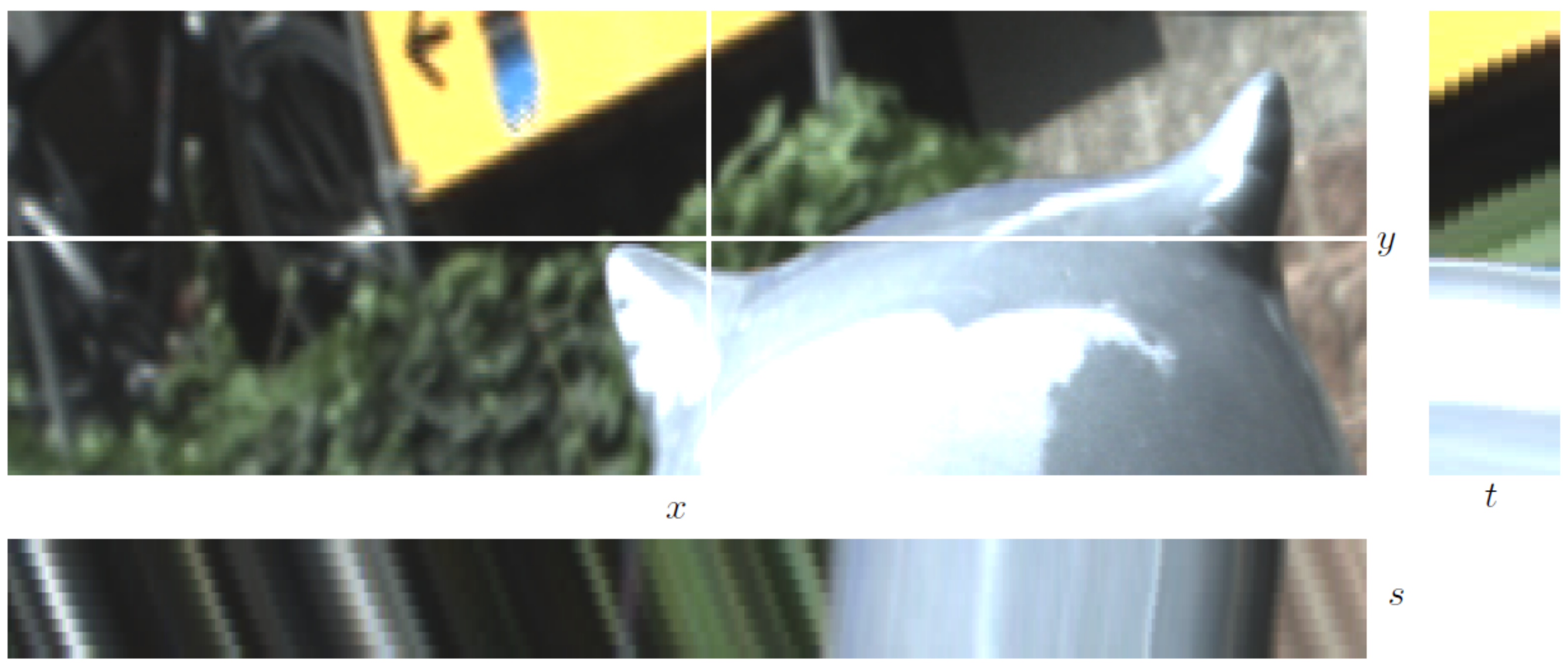
Contributions

- End-to-end deep neural network for non-Lambertian intrinsic light field decomposition together with disparity estimation
- Skip connections from the encoder to corresponding decoder parts
- Sequence of 2D convolutions acting on the spatial and angular domains instead of "heavy" 3D convolutions
- Combination of supervised and unsupervised training

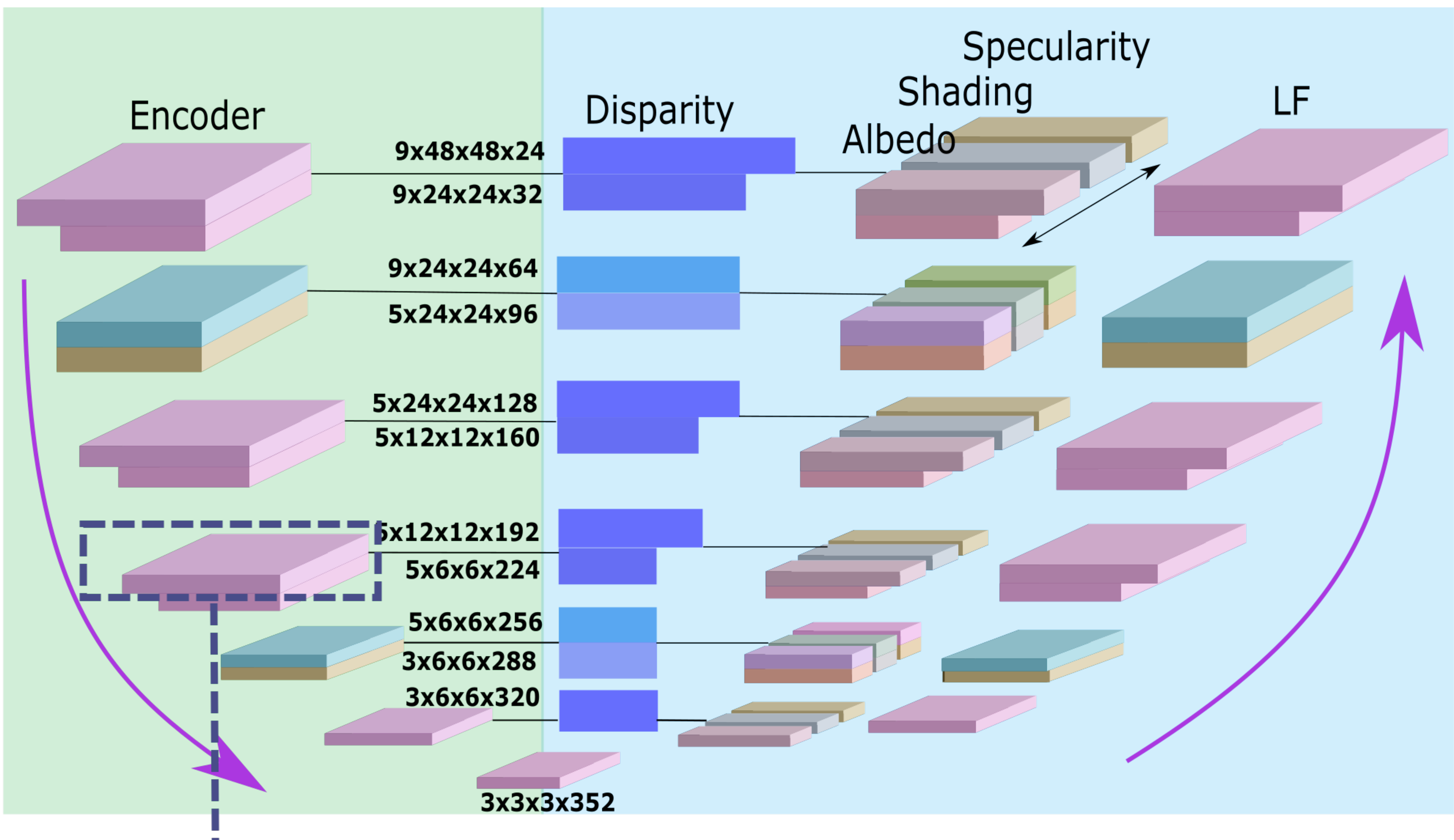


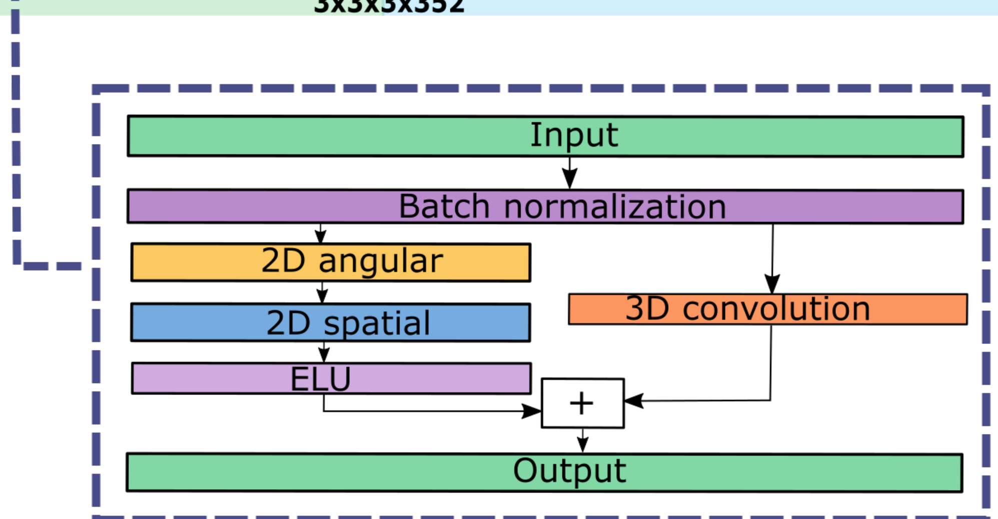
Light field structure

Defined on 4D ray space $\mathcal{R} = \Pi \times \Omega$, which parametrizes rays $\boldsymbol{r} = (x, y, s, t)$ by their intersection coordinates with two planes Π and Ω, [4]. Intersection with the focal plane Π gives view point coordinates (s, t), while the image plane Ω denotes image coordinates (x, y).



Network architecture





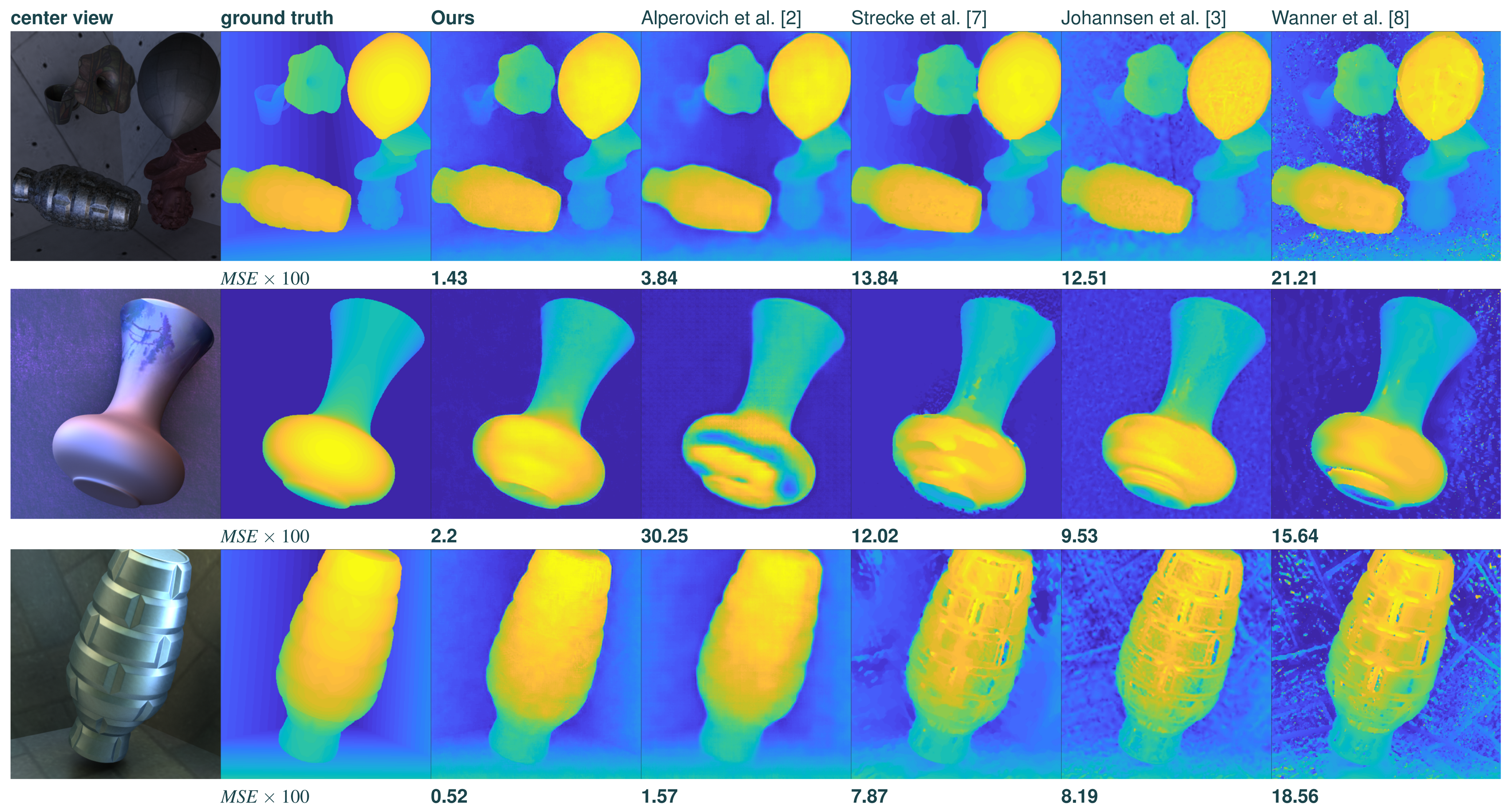
Inputs:
- a pair of $9 \times 96 \times 96 \times 3$ horizontal and vertical 3D slices of the light field

Structure:
- 12 residual blocks in the encoding and four decoding pathways
- encoder features are copied as the skip connections to the decoders
- intrinsic components modeled as 3D decoders, while disparity is 2D
- 3D convolutions are replaced with the sequence of 2D convolutions acting on the angular and spatial domains

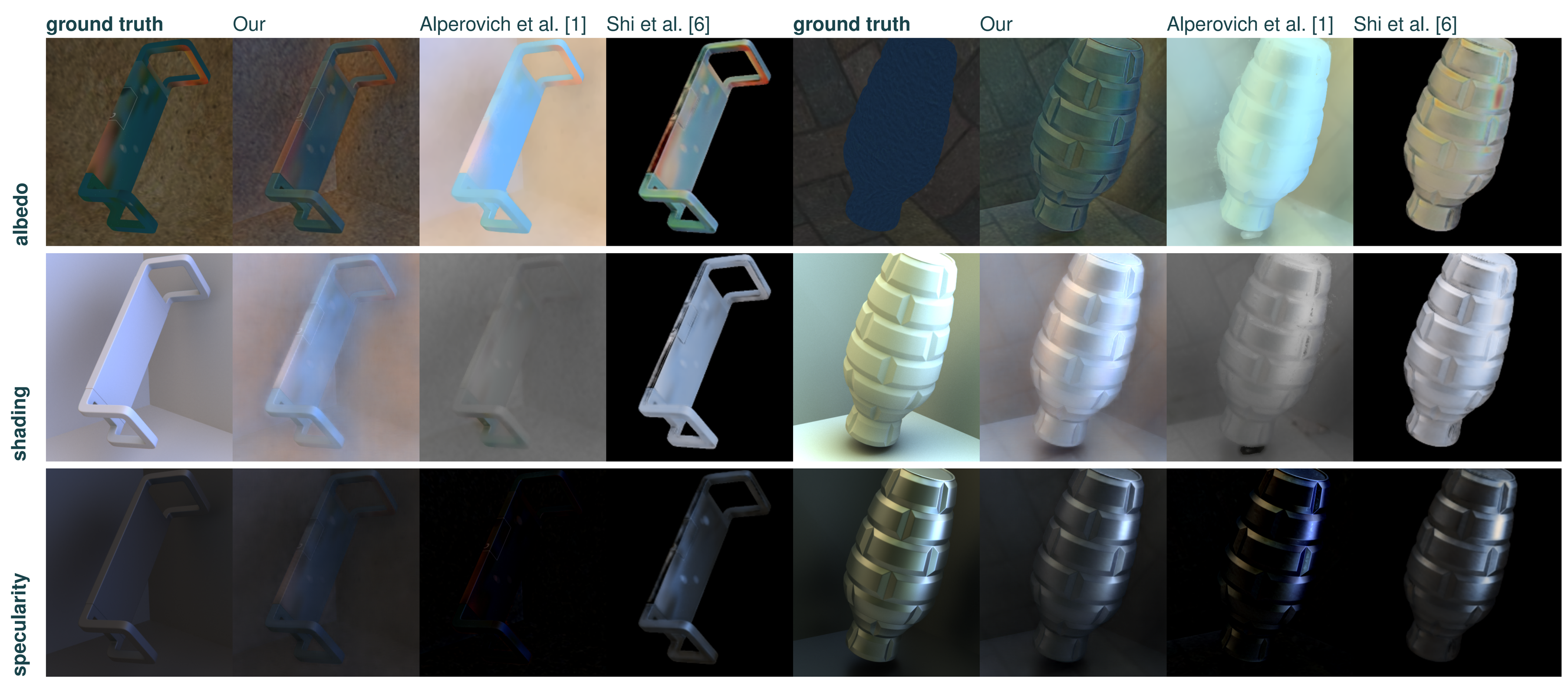
Loss:
- scale invariant MSE for albedo and shading
- standard MSE for specular component and disparity
- dichromatic reflection model [5]

This work was supported by the ERC Starting Grant "Light Field Imaging and Analysis" (LIA 336978, FP7-2014).
Presented at the European Signal Processing Conference (EUSIPCO), Rome, September 2018.

Synthetic results



Ground truth and estimated disparity for three synthetic data sets generated with Blender. The disparity range is $[-2.12, 2.51]$. We compare with the recent deep network [2] that jointly decomposes input light field into diffuse and specular components and estimates the disparity and three modeling approaches [3], [7] and [8] which are based on dictionary learning, focal stack symmetry and orientation of EPI patches.



Comparison on two synthetic data sets generated with Blender. The light field size is $9 \times 9 \times 512 \times 512 \times 3$. We compare with the modeling approach for light fields [1] and single image CNN [6].

Real world results

References

[1] A. Alperovich, O. Johannsen, M. Strecke, and B. Goldluecke. Shadow and specularity priors for intrinsic light field decomposition. In *EMMCVPR*, 2017.
[2] A. Alperovich, O. Johannsen, M. Strecke, and B. Goldluecke. Light field intrinsics with a deep encoder-decoder network. In *Proc. CVPR*, 2018.
[3] O. Johannsen, A. Sulc, and B. Goldluecke. What sparse light field coding reveals about scene structure. In *Proc. CVPR*, 2016.
[4] M. Levoy. Light fields and computational imaging. *Computer*, 39(8):46–55, 2006.
[5] Steven Shafer. Using color to separate reflection components. *Color Research & Application*, 10(4):210–218, 1985.
[6] J. Shi, Y. Dong, H. Su, and S. Yu. Learning non-lambertian object intrinsics across shapenet categories. In *Proc. CVPR*, 2017.
[7] M. Strecke, A. Alperovich, and B. Goldluecke. Accurate depth and normal maps from occlusion-aware focal stack symmetry. In *Proc. CVPR*, 2017.
[8] S. Wanner and B. Goldluecke. Globally consistent depth labeling of 4D light fields. In *Proc. CVPR*, pages 41–48, 2012.