Light-Field Intrinsic Dataset

Sumit Shekhar *[1]
sshekhar@mpi-inf.mpg.de
Shida Kunz Beigpour *[1]
shida@mpi-inf.mpg.de
Matthias Zeigler[2]
matthias.ziegler@iis.fraunhofer.de
Michał Chwesiuk[3]
mchwesiuk@wi.zut.edu.pl
Dawid Paleń[3]
dawid60@gmail.com
Karol Myszkowski[1]
http://people.mpi-inf.mpg.de/~karol/
Joachim Keinert[2]
joachim.keinert@iis.fraunhofer.de
Radosław Mantiuk[3]
http://rmantiuk.zut.edu.pl/
Piotr Didyk[1,4]
https://people.mpi-inf.mpg.de/~pdidyk/

[1] Max Planck Institute for Informatics
Saarland Informatics Campus
Saarbrücken, Germany

[2] Fraunhofer Institute for Integrated Circuits IIS
Erlangen, Germany

[3] West Pomeranian University of Technology
Szczecin, Poland

[4] Università della Svizzera italiana
Lugano, Switzerland

Abstract

Light-field imaging has various advantages over the traditional 2D photography, such as depth estimation and occlusion detection, which can aid intrinsic decomposition. The extracted intrinsic layers enable multiple applications, such as light-field appearance editing. However, the current light-field intrinsic decomposition techniques primarily resort to qualitative comparisons, due to lack of ground-truth data. In this work, we address this problem by providing intrinsic dataset for real world and synthetic 4D and 3D (only horizontal parallax) light fields. The ground-truth intrinsic data comprises albedo, shading and specularity layers for all sub-aperture images. In case of synthetic data, we also provide ground-truth depth, normals, and further decomposition of shading into direct and indirect components. For real-world data acquisition, we make use of custom hardware and 3D printed objects, assuring precision during multi-pass capturing. We also perform, qualitative and quantitative, comparison of existing intrinsic decomposition algorithms for single image, video, and light field. To the best of our knowledge, this is the first such dataset for light fields, which is also applicable for single image, multi-view stereo, and video.

© 2018. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

*indicates equal contribution

1 Introduction

An image captured with a conventional camera is a result of a complex process that depends on camera settings as well as scene configuration. Such scene parameters comprise scene geometry, illumination, object material, participating media, viewing direction, and other properties. In the field of computer vision, decomposing an image into its *intrinsic* properties is vital for scene understanding [14]. Intrinsic decomposition can enable applications like recoloring, object segmentation, compositing, appearance editing, etc. One can aim to extract different intrinsic properties from a given image, such as reflectance [13, 33, 38, 43] or geometry [25, 44, 49, 53]. Extraction of specularity intrinsic layer is a problem on its own [9]. Many algorithms have been proposed for extracting intrinsic layers from single images [19, 54], single images with depth [11, 24, 29], multi-views [35, 51], videos [21, 34, 41, 52] and light fields [6, 7, 8, 27, 46]. Recently, some methods have been developed that use machine learning techniques for intrinsic decomposition [15, 30, 42, 45, 55].

In most of the above-mentioned methods, authors perform a qualitative comparison with others or a quantitative evaluation with respect to synthetic data. A lack of enough real-world intrinsic ground-truth datasets is the reason for such approach. Moreover the data-driven techniques cannot generalize easily considering lack of training data, especially for real world scenes.

A light field aims to record the light information in a given volume or passing through a plane. The light-field technology is gaining more and more attention due to the popularity of virtual- and augmented-reality. Adelson and Bergen [4] gave a generic definition of light field, which considers all rays in a volume. In our work, we restrict ourselves to the two-plane parametrization of light field [37].

Using light fields leverages the task of intrinsic decomposition, since depth reconstruction and occlusion detection benefit from the high number of available views. Moreover extracting intrinsic layers also enables various applications, like light-field appearance editing. However one of the major difficulties in dealing with real-world wide-baseline light fields, lies in its capturing process. It is mandatory to calibrate incorporated cameras with high precision. On top of this, acquisition of light-field intrinsic layers requires multiple capturing passes that need to be aligned precisely in terms of both camera and object position.

Our contribution consists of two main parts. Firstly, we provide real-world ground-truth intrinsic layers for 3D and 4D light fields. Secondly, we render synthetic ground-truth intrinsic layers for 4D light fields. In case of synthetic data, we also provide ground-truth depth, normal and further decomposition of shading into *direct* and *indirect* components.

2 Related Work

In this section, we review existing ground truth intrinsic datasets. Then, we discuss challenges associated with wide-baseline light-field acquisition and existing capturing setups. Finally, we discuss existing intrinsic decomposition algorithms with an emphasis on those which can handle light-field data.

Ground Truth Intrinsic Data The work of Grosse *et al.*,[28] (also known as the *MIT dataset*) is the first real world intrinsic decomposition dataset. Later, Barron *et al.*,[10, 12] rendered MIT dataset models using natural illumination. Beigpour *et al.*,[16, 17] improved upon the real-world MIT dataset by using multi-illuminants and also considering multiple views in the later work. The MPI *Sintel* dataset [23, 50] has become popular as a synthetic scene benchmark for intrinsic decomposition evaluation. However, the shading layer, in *Sintel* dataset is obtained by dividing the *clean pass* of the scene (containing non-lambertian objects) with the *albedo pass*, which is physically incorrect. We consider a separate specularity layer to handle such scenario. In a recent work, Bonneel *et al.*,[22] provide synthetic intrinsic ground-truth data for limited number of images. Shi *et al.*,[45] provide ground-truth rendering data for 3D objects from *ShapeNet* database. To the best of our knowledge, no dataset exist for light fields providing all intrinsic layers of shading, albedo, and specularity.

Wide-Baseline Light-Field Acquisition The parallax of a light-field dataset is one of the most important properties when processing light-field data. As a rule of thumb, we denote datasets with stereo parallax $\leq 1px$ as *dense* (or *narrow-baseline*) light field. Datasets that exceed $1px$ are denoted as *sparse* (or *wide-baseline*) light field. The wide-baseline dense camera arrays by Vaish *et al.*, [48] was one of the first attempt to capture light fields with wide-baseline. Kim *et al.*, [31] used a setup with 1.5 m slider for denser sampling of several outdoor scenes. Different types of camera array setups have been proposed in the *Stanford Light Field Archive* [3]. Their latest setup was a two dimensional system build from Lego bricks carrying a DSLR camera.

Adhikarla *et al.*, [5] develop a one-meter long motorized linear stage for capturing real-world scenes. Ziegler *et al.*,[56] use cantilever axes to capture natural scene spreading in the order of meters in both horizontal and vertical direction. One of the important challenge in all the above setups is camera calibration and precision of camera re-positioning. We use the setup similar to Ziegler *et al.*, [56] for real-world 4D light-field capturing, and for 3D light field (only horizontal parallax), we employ the setup similar to Adhikarla *et al.*, [5].

Intrinsic Image Decomposition A broad survey of intrinsic image decomposition algorithms is presented by Bonneel *et al.*, [22] and Ma *et al.*, [40]. The *retinex theory* of color vision, introduced by Land [36], formed the basis of many intrinsic decomposition algorithms. Hachama *et al.*,[29] and Chen *et al.*,[24] use RGB-D images to include priors based on normals. Xie *et al.*,[51] employ the disparity information from the given multi-view stereo data to introduce additional constraints. Meka *et al.*,[41] use a combination of local and global spatio-temporal priors to achieve real-time performance. Duchêne *et al.*,[26] consider lighting conditions along with shadows to compute intrinsic layers with an application towards relighting. Tunwattanapong *et al.*, [47] use a rotating arc of LEDs (spherical gradient illumination) to extract reflections and normals in world-space for both diffuse and specular components. Kim *et al.*, [32], use a two-way polarized light-field (TPLF) camera to show layered reflectance separation in the angular domain for human faces. The *Direct Intrinsic* from Narihara *et al.*,[42] was one of the first data-driven approach to solve intrinsic decomposition problem. Later, Shi *et al.*,[45] also considered non-lambertian objects in their training data. Baslamisli *et al.*, [15] incorporate traditional intrinsic decomposition priors in their custom loss function. Garces *et al.*,[27] and Alperovich *et al.*,[6] focus on intrinsic decomposition of light fields with narrow baseline. In a follow-up work Alperovich *et al.*,[7] introduced priors to tackle cast shadows and inter-reflections. In a recent work Alperovich *et al.*, [8] use an encoder-decoder network to extract specular and diffuse components for narrow-baseline light fields. In another recent work, specifically aimed for wide-baseline light-fields, Beigpour *et al.*,[18] take inspiration from multi-view stereo intrinsic decomposition to get consistent results.

3 Data Acquisition

In this section we describe the setup and capturing details for real scenes and the rendering specifications for the synthetic data. Our real-world acquisition step is inspired by the work of Beigpour *et al.*, [16] and Grosse *et al.*, [28]. However, we extend it for light fields and use only 3D printed objects. In order to obtain synthetic ground-truth data we make use of physically based rendering in *Blender (Cycles)* [20]. Our ground-truth data consists of *shading* (comprising interaction of scene illumination with geometry), *albedo* (diffuse scene reflective component, independent of view direction), and *specularity* (directional reflectance component) intrinsic layers for each sub-aperture image of the light field.

Image Formation Model As defined by Levoy and Hanrahan [37], a light field can be considered as a collection of images. We, thus, generalize the image formation model used by Grosse *et al.*,[28] for light fields. We assume that each sub-aperture image (I) in a light field is composed of diffuse (I_d) and specular (C) components.

$$I = I_d + C \tag{1}$$

The diffuse component can be further expressed as the product of shading (S) and albedo (A) layers.

$$I_d = S \cdot A \tag{2}$$

Our ground-truth extraction methodology of intrinsic layers is similar to the work of Beigpour *et al.*, [16]. The extracted shading ($\tilde{S}$) and albedo ($\tilde{A}$) are relative and proportional to the absolute values of S and A respectively,

$$\tilde{S} \propto S \quad \text{and} \quad \tilde{A} \propto A \tag{3}$$

The relative values of albedo and shading are extracted in a way such that the product equals I_d. For brevity, we omit the pixel co-ordinate $\mathbf{x}$. However, Eqs. 1-3 hold for all pixels in all linear-encoded sub-aperture images.

3.1 Capturing Real-World Intrinsic Layers

The real-world dataset consists of intrinsic layers for three 4D and three 3D light fields. The scene setup, for both cases, consists of 3D printed objects arranged on a horizontal platform, see Fig. 1b. We use two flicker-free LED based lamps for scene illumination. The lamps are DC powered in order to maintain constant illumination during the capturing process. In order to capture the scenes we have used Canon EOS 6D and 5D as well as Sony Alpha 7 R II cameras. All the cameras were equipped with high-resolution full-frame imaging sensors and a 50 mm or 28 mm Canon lens, respectively. Refer to supplementary material for specific camera and lens configuration used in each scenario. The camera plane is perpendicular to the scene layout as shown in Fig. 1b. We cover the area surrounding the scene with diffuse black cloth in order to minimize inter-reflections and ambient lighting.

3D Light-Field Capturing: In order to capture 3D light field, we follow the approach of Adhikarla *et al.*,[5]. A high-quality camera is mounted on a linear motor stage. The camera movement is controlled using a stepper-motor and an Arduino board (as shown in

(a) 3D models **(b) Scene Layout**

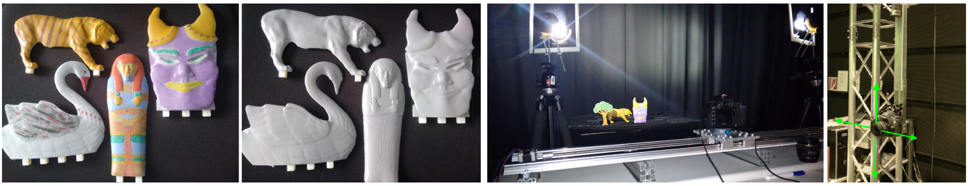

Figure 1: (a) The two set of 3D printed objects, colored and gray. (b) Light-field capturing setups for (left) 3D (linear motor stage with a mounted camera and light-sources with polarizing filters) and (right) 4D (camera mounted on a cantilever axes moving along the directions shown in the figure) light fields.

Fig. 1b). The difference between consecutive camera positions is small leading to dense light-field acquisition. The precision in camera movement is in the order of μm. Please refer to Adhikarla *et al.*,[5] for more details. The whole capturing system is relatively mobile and can be easily installed on top of a study-table (see Fig. 1b).

4D Light-Field Capturing: We take inspiration from Ziegler *et al.*,[56] for capturing wide-baseline light fields. The camera moves both in horizontal and vertical directions, along the camera plane (indicated by green arrows in Fig. 1b). The camera can be translated by up to 4 m horizontally and 0.5 m in vertical direction with a precision error of 80 μm. Please refer to Ziegler *et al.*,[56] for further details. In contrast to the *3D Light-Field Capturing*, this system consists of large mechanical moving parts, and the whole setup is fixed thereby posing restrictions in terms of scene setting.

In both setups mentioned above, the large translation of the camera enables capturing light fields with large parallax.

Specularity Extraction: We use the approach introduced in Grosse *et al.*,[28] to capture the specularity layer for sub-aperture images. The colored objects are placed on the horizontal platform as shown in Fig. 1b. We mount a linear polarizing filter in front of each of the light sources, so that the light illuminating the scene is polarized. Another polarizing filter is mounted on the camera lens. The specular and diffuse versions of the scene are captured in two runs. In the first run, we tune the polarizing filter on the camera so that the specular and diffuse reflection passes the filter. In the second run, we tune the filter to block the specular reflection. Thus, we capture each sub-aperture image (I) and its corresponding diffuse version (I_d), respectively. The specularity layer (C) is obtained using Eq. 1. The extension of single image based specularity extraction for light fields is not 100% accurate. However the inaccuracies are negligible, we further discuss this as a limitation in Sec. 5.

Albedo and Shading Extraction: In order to capture the shading version of the scene, we use a second, identical, set of objects painted with diffuse gray color (RAL7042 paint: [R-142, G-146, B-145]). These objects should align precisely with their colored counterpart at pixel level. We ensure such alignment accuracy by using two sets of 3D printed objects and tight fitting rigid pins for their placement. The two set of objects are printed using a high precision printer [1] for the above requirement (see Fig. 1a), an alternative approach of using just one set of objects and painting it twice, as done by Beigpour *et al.*,[16], would destroy the objects original color and texture making it impossible to repeat the acquisition step is needed (e.g., due to errors or to test different variations of the same scene).

The ambient illumination in the scene is suppressed by the black surrounding curtains. Once the objects are placed properly, the polarizing filter is tuned such that only diffuse reflection can pass. Thus the relative shading ($\widetilde{S}$), diffuse gray version of the scene, is captured. The relative albedo ($\widetilde{A}$), diffuse reflectance, is computed using Eq. 2. Fig. 2 shows real-world scenes captured for 3D and 4D light fields. In Fig. 3, we show the ground truth intrinsic layers for two scenes, *"3D Scene 2"* and *"4D Scene 2"* respectively.

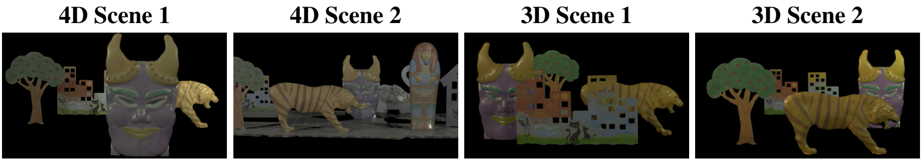

Figure 2: Central view of the two out of three captured real-world 3D and 4D light fields.

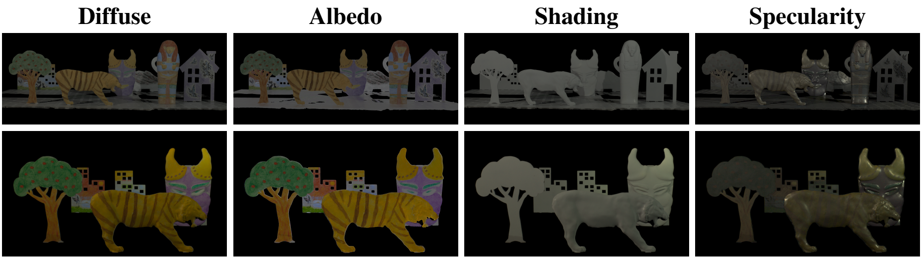

Figure 3: Ground-truth real-world intrinsic layers for *4D Scene 2* (top) and *3D Scene 2* (bottom).

3.2 Rendering Intrinsic Layers

As compared to real-world capturing, we have more freedom in terms of scene selection and rendering of multiple intrinsic properties for synthetic light fields. We use *Cycles* integrated in *Blender* for physically based rendering. We have used open-source scenes available in *Blend Swap* [2] with minor modifications. The blender scenes are selected in a manner such that they look realistic, are difficult for intrinsic decomposition, and cover a wide variety of reflectance textures. The light field is rendered by regular rectangular sampling of the camera plane (which is perpendicular to viewing direction). For all sub-aperture images, we also render the intrinsic layers of albedo (A), depth, normal, direct (S_d), and indirect (S_i) shading.

Albedo, Shading, and Specularity Extraction: As stated previously, albedo is directly rendered for each view. Shading is obtained as the sum of its direct and indirect components,

$$S = S_d + S_i \tag{4}$$

The direct shading (S_d) is the part of shading caused by *direct illumination* (single bounce of rays from the scene objects), and indirect shading (S_i) is formed due to *indirect illumination* (multiple bounces of the rays from the scene objects). The diffuse version of the view is obtained by multiplying albedo and shading as in Eq. 2. The specular layer is obtained as the difference between the original image and its diffuse version using Eq. 1. In Fig. 4, we show the center view, and in Table 3.2 we depict the parallax values for all the rendered 4D light fields. The intrinsic layers, rendered for the center view, of *Scene 2* are depicted in Fig. 5. In *Scene 7*, we consider the interesting case of human skin modeling. Since the skin reflectance cannot be explained by our image formation model (see Eq. 1), we further decompose the *non-specular* component into *only diffuse* and *subsurface scattering* components, see Fig. 6, similar to Kim *et al.*,[32].

We commit to provide modified open-source scenes, scripts for distributed rendering (over a cluster) and post-processing. Note that all images shown in this work are gamma corrected and scaled for better visualization. The provided dataset contains both gamma corrected images and images with linearly encoded pixel values.

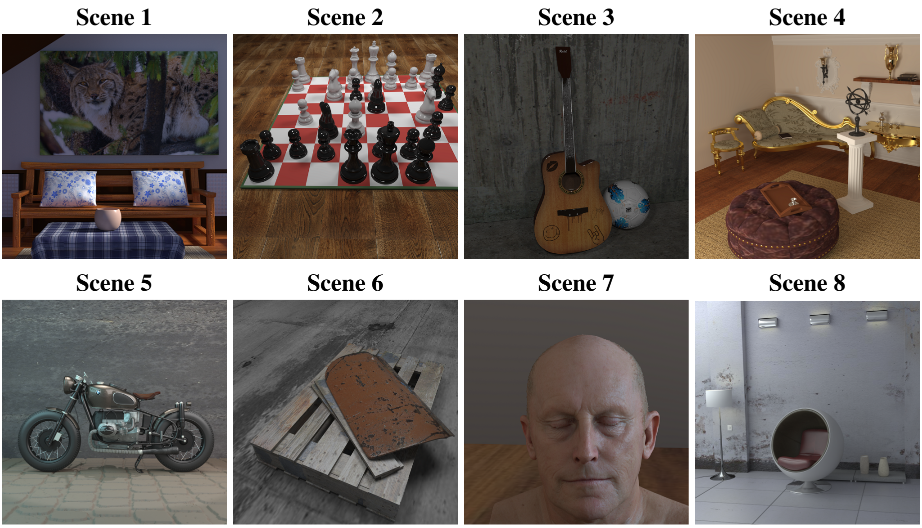

Figure 4: Central view of all the synthetic rendered scenes.

4 Evaluation of The Results

We evaluate four intrinsic decomposition algorithms, namely Bell *et al.*,[19] for single images, Meka *et al.*,[41] for videos, Alperovich *et al.*,[7] and Garces *et al.*,[27] for light fields using our dataset. All the results presented in this paper, for different algorithms, are based on respective author implementation and for Garces *et al.*,we only perform qualitative comparison. We calculate the average error per sub-aperture image for each of these algorithms using *DSSIM* [39] as an error metric, $DSSIM(x,y) = (1 - SSIM(x,y))/2$, which better corresponds to the human perception than MAE or RMSE.

We have used a Matlab implementation for calculating *SSIM*(Structural Similarity Index) values. In order to measure consistency of intrinsic decomposition between views, we

Scene Number	Min. Depth (in m)	Max. Depth (in m)	Stereo Parallax (in pixels)
1	2.44	4.39	17.41
2	10.29	49.07	22.03
3	10.59	17.67	21.71
4	2.13	4.94	25.54
5	6.15	15.92	44.19
6	4.45	16.33	12.16
7	11.69	26.85	12.32
8	3.81	6.51	17.66

Table 1: Minimum and maximum object distances and parallax between extreme views in one direction for our dense light-field rendering.

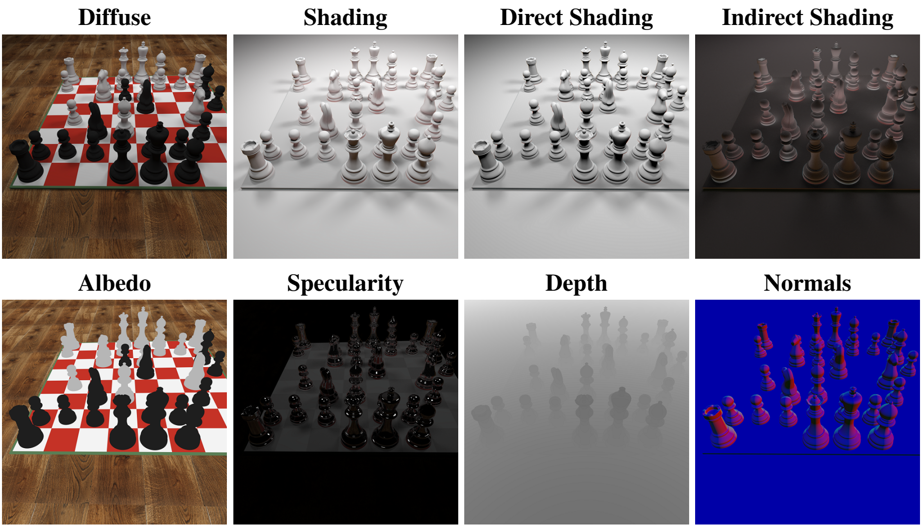

Figure 5: The rendered intrinsic layers for scene 2.

Figure 6: The non-specular component in case of a face scene can be further decomposed into *only diffuse* and *subsurface* scattering components.

compute the variance in error for all sub-aperture images, see Table 4, comparing different intrinsic decomposition methods. We show qualitative comparison of intrinsic decomposition and specularity extraction algorithms in Fig. 7 and Fig. 8 respectively.

Scene Type/Num	Single Image (Bell *et al.*,)		Video (Meka *et al.*,)		Light Field (Alperovich *et al.*,)	
	μ	σ	μ	σ	μ	σ
Syn./ 2	$1.05e-1$	$4.66e-4$	$1.99e-1$	$2.33e-8$	$1.56e-1$	$4.20e-7$
Syn./ 3	$2.19e-1$	$1.48e-3$	$2.87e-1$	$5.95e-8$	$8.20e-2$	$9.22e-6$
Real 3D/ 2	$4.92e-2$	$8.83e-5$	$8.79e-2$	$2.01e-6$	NA	NA
Real 3D/ 3	$3.10e-2$	$4.22e-6$	$7.31e-2$	$1.20e-6$	NA	NA
Real 4D/ 2	$4.23e-2$	$3.88e-5$	$3.69e-2$	$1.02e-6$	$4.67e-1$	$2.34e-5$
Real 4D/ 3	$4.38e-2$	$5.59e-5$	$3.69e-2$	$1.75e-6$	$4.66e-1$	$2.01e-5$

Table 2: The average (μ) and variance (σ) of error in *albedo* extraction, for a set of sub-aperture images for *real* and *synthetic* data. Refer to supplementary for a similar table for *shading* extraction.

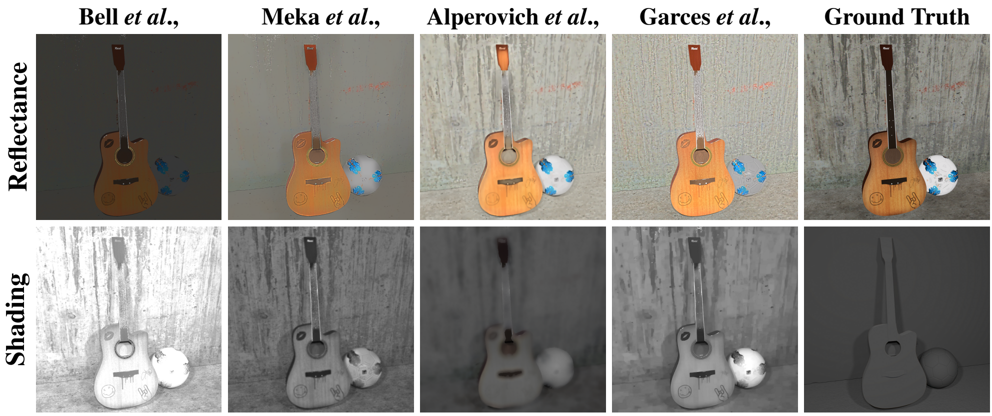

Figure 7: A comparison of intrinsic decomposition methods for single image (Bell *et al.*,[19]), video (Meka *et al.*,[41]) and light fields (Alperovich *et al.*,[7] and Garces *et al.*,[27]) using our ground-truth data.

Figure 8: A comparison of specularity extraction methods for light fields.

5 Discussion

Applications: Though supporting research on intrinsic decomposition methods is our main contribution, we want to emphasize that our rich dataset can be used for other applications as well. The dense, real-world 3D and synthetic 4D, light fields, can be used to evaluate light-field reconstruction or depth-image-based rendering (DIBR). For the 3D printed objects, we provide corresponding geometry data. In case of synthetic images, we render depth and normal passes. The above two points make our dataset suitable for applications like *shape from shading* and *3D Reconstruction*. In our dataset, we provide a separate specular layer. Thus making our data suitable for assessing *specularity removal* techniques. For our synthetic rendering, we consider *indirect* shading pass which can be used for judging algorithms that aim to compute *indirect* illumination.

A light field can be considered as a special case of *multi-view stereo*, comprising of multiple views (*single images*), which can also be visualized in *video* format. Thus our dataset is applicable for different data modalities.

Limitations: Due to varying viewing angle, the above methodology of real-world specularity extraction is not completely accurate. In an ideal scenario, one would use only distant light source with a linear polarizing filter. Otherwise the orientation of the polarizing filter (in front of the camera) needs to be re-adjusted every time the camera position changes. However, in practice these inaccuracies, in terms of specularity leakage in diffuse pass, were not noticeable visibly.

6 Conclusion

We provide intrinsic dataset for real and synthetic light fields. We believe that our rich dataset will contribute to the research of light-field intrinsic decomposition and other applications, as discussed in Sec. 5. Our dataset contains challenging surface color texture, complex geometry, and moving specularities. In case of real-world data acquisition, we make use of custom hardware and 3D printed objects, to ensure precise alignment for our multi-pass capturing scenario. We also provide scripts to render (using a cluster) and process light-field intrinsic layers, thereby enabling others to render such layers for an arbitrary scene. We believe that this is the only light-field dataset that provides ground-truth intrinsic layers of albedo, shading, and specularity.

7 Acknowledgements

We would like to thank Michal Piovarči for his help with the 3D printing. We would like to thank Anna Alperovich and Ole Johansen for valuable discussion. We would like to thank Anna Alperovich, Abhimitra Meka, and Elena Garces for kindly providing the necessary comparisons. We thank the reviewers for their insightful comments. The project was supported by the Fraunhofer and Max Planck cooperation program within the German pact for research and innovation (PFI).

References

[1] Objet260 connex3, stratasys. http://www.stratasys.com/de/3d-printers/design-series/objet260-connex3. Accessed: 2018-07-19.

[2] Blend swap. https://www.blendswap.com/. Accessed: 2018-04-23.

[3] The (new) stanford light field archive, 2008. URL http://lightfield.stanford.edu/index.html.

[4] Edward H Adelson, James R Bergen, et al. The plenoptic function and the elements of early vision. 1991.

[5] Vamsi Kiran Adhikarla, Marek Vinkler, Denis Sumin, Rafał Mantiuk, Karol Myszkowski, Hans-Peter Seidel, and Piotr Didyk. Towards a quality metric for dense light fields. In *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[6] A. Alperovich and B. Goldluecke. A variational model for intrinsic light field decomposition. In *Asian Conference on Computer Vision (ACCV)*, 2016.

[7] A. Alperovich, O. Johannsen, M. Strecke, and B. Goldluecke. Shadow and specularity priors for intrinsic light field decomposition. In *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, 2017.

[8] A. Alperovich, O. Johannsen, M. Strecke, and B. Goldluecke. Light field intrinsics with a deep encoder-decoder network. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[9] Alessandro Artusi, Francesco Banterle, and Dmitry Chetverikov. A survey of specularity removal methods. In *Computer Graphics Forum*, volume 30, pages 2208–2230, 2011.

[10] J. T. Barron and J. Malik. Intrinsic scene properties from a single rgb-d image. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 17–24, June 2013.

[11] Jonathan Barron and Jitendra Malik. Intrinsic scene properties from a single RGB-D image. In *IEEE CVPR*, 2013.

[12] Jonathan T. Barron and Jitendra Malik. Color constancy, intrinsic images, and shape estimation. In *Proceedings of the 12th European Conference on Computer Vision - Volume Part IV*, ECCV'12, pages 57–70, Berlin, Heidelberg, 2012. Springer-Verlag. ISBN 978-3-642-33764-2.

[13] Jonathan T Barron and Jitendra Malik. Shape, illumination, and reflectance from shading. *IEEE transactions on pattern analysis and machine intelligence*, 37(8):1670–1687, 2015.

[14] H. G. Barrow and J. M. Tenenbaum. *Recovering Intrinsic Scene Characteristics from Images*. Academic Press, 1978.

[15] Anil S. Baslamisli, Hoang-An Le, and Theo Gevers. Cnn based learning using reflection and retinex models for intrinsic image decomposition. *arXiv preprint arXiv:1712.01056*, 2017.

[16] Shida Beigpour, Andreas Kolb, and Sven Kunz. A comprehensive multi-illuminant dataset for benchmarking of intrinsic image algorithms. In *Proc. IEEE International Conference on Computer Vision (ICCV)*, pages 172–180, December 2015.

[17] Shida Beigpour, Mai Lan Ha, Sven Kunz, Andreas Kolb, and Volker Blanz. Multi-view multi-illuminant intrinsic dataset. In *Proc. British Machine Vision Conference (BMVC)*, 2016.

[18] Shida Beigpour, Sumit Shekhar, Mohsen Mansouryar, Karol Myszkowski, and Hans-Peter Seidel. Light-field appearance editing based on intrinsic decomposition. *Journal of Perceptual Imaging (JPI)*, 2018.

[19] Sean Bell, Kavita Bala, and Noah Snavely. Intrinsic images in the wild. *ACM Trans. on Graphics*, 33(4), 2014.

[20] Blender Foundation. Blender. URL https://www.blender.org/.

[21] Nicolas Bonneel, Kalyan Sunkavalli, James Tompkin, Deqing Sun, Sylvain Paris, and Hanspeter Pfister. Interactive intrinsic video editing. *ACM Trans. on Graphics*, 33(6): 197, 2014.

[22] Nicolas Bonneel, Balazs Kovacs, Sylvain Paris, and Kavita Bala. Intrinsic Decompositions for Image Editing. *Computer Graphics Forum*, 2017. ISSN 1467-8659. doi: org:443/handle/10.1111/cgf13149.

[23] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In A. Fitzgibbon et al. (Eds.), editor, *European Conf. on Computer Vision (ECCV)*, Part IV, LNCS 7577, pages 611–625. Springer-Verlag, October 2012.

[24] Qifeng Chen and Vladlen Koltun. A simple model for intrinsic image decomposition with depth cues. In *Proc. of IEEE Intl. Conference on Computer Vision*, pages 241–248, 2013.

[25] Angela Dai, Matthias Nießner, Michael Zollöfer, Shahram Izadi, and Christian Theobalt. BundleFusion: Real-time Globally Consistent 3D Reconstruction using On-the-fly Surface Re-integration. *arXiv preprint arXiv:1604.01093*, 2016.

[26] Sylvain Duchêne, Clement Riant, Gaurav Chaurasia, Jorge Lopez-Moreno, Pierre-Yves Laffont, Stefan Popov, Adrien Bousseau, and George Drettakis. Multi-view intrinsic images of outdoors scenes with an application to relighting. *ACM Transactions on Graphics*, 2015.

[27] Elena Garces, Jose I Echevarria, Wen Zhang, Hongzhi Wu, Kun Zhou, and Diego Gutierrez. Intrinsic light field images. *Computer Graphics Forum*, 36(3), 2017.

[28] R. Grosse, M.K. Johnson, E.H. Adelson, and W.T. Freeman. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *IEEE ICCV*, pages 2335–2342, 2009.

[29] Mohamed Hachama, Bernard Ghanem, and Peter Wonka. Intrinsic scene decomposition from RGB-D images. In *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*, pages 810–818, 2015.

[30] Michael Janner, Jiajun Wu, Tejas Kulkarni, Ilker Yildirim, and Joshua B Tenenbaum. Self-Supervised Intrinsic Image Decomposition. In *Advances In Neural Information Processing Systems*, 2017.

[31] Changil Kim, Henning Zimmer, Yael Pritch, Alexander Sorkine-Hornung, and Markus Gross. Scene reconstruction from high spatio-angular resolution light fields. *ACM Trans. Graph.*, 32(4):73:1–73:12, July 2013. ISSN 0730-0301.

[32] Jaewon Kim, Shahram Izadi, and Abhijeet Ghosh. Single-shot layered reflectance separation using a polarized light field camera. In Elmar Eisemann and Eugene Fiume, editors, *Eurographics Symposium on Rendering - Experimental Ideas & Implementations*. The Eurographics Association, 2016. ISBN 978-3-03868-019-2. doi: 10.2312/sre.20161204.

[33] Kihwan Kim, Jinwei Gu, Stephen Tyree, Pavlo Molchanov, Matthias Nießner, and Jan Kautz. A lightweight approach for on-the-fly reflectance estimation. In *ICCV*, pages 20–28. IEEE Computer Society, 2017.

[34] Pierre-Yves Laffont and Jean-Charles Bazin. Intrinsic decomposition of image sequences from local temporal variations. In *The IEEE International Conference on Computer Vision (ICCV)*, 2015.

[35] Pierre-Yves Laffont, Adrien Bousseau, Sylvain Paris, Frédo Durand, and George Drettakis. Coherent intrinsic images from photo collections. *ACM Trans. on Graphics (Proc. SIGGRAPH Asia)*, 31(6), 2012.

[36] E.H. Land. *The Retinex Theory of Color Vision*. Scientific American offprints. W.H. Freeman Company.

[37] Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, 1996.

[38] Stephen Lombardi and Ko Nishino. Radiometric scene decomposition: Scene reflectance, illumination, and geometry from RGB-D images. *CoRR*, abs/1604.01354, 2016.

[39] Artur Loza, Lyudmila Mihaylova, Nishan Canagarajah, and David Bull. Structural similarity-based object tracking in video sequences. In *International Conference on Information Fusion*, pages 1–6. IEEE, 2006.

[40] Yupeng Ma, Xiaoyi Feng, Xiaoyue Jiang, Zhaoqiang Xia, and Jinye Peng. Intrinsic image decomposition: A comprehensive review. pages 626–638, 12 2017.

[41] Abhimitra Meka, Michael Zollhöfer, Christian Richardt, and Christian Theobalt. Live intrinsic video. *ACM Trans. on Graphics (Proceedings SIGGRAPH)*, 35(4), 2016.

[42] Takuya Narihira, Michael Maire, and Stella X. Yu. Direct intrinsics: Learning albedo-shading decomposition by convolutional regression. In *International Conference on Computer Vision (ICCV)*, 2015.

[43] Geoffrey Oxholm and Ko Nishino. Shape and reflectance estimation in the wild. *IEEE Trans. Pattern Anal. Mach. Intell.*, 38(2):376–389, 2016.

[44] Yvain Quéau, Jean Mélou, Jean-Denis Durou, and Daniel Cremers. Dense multi-view 3d-reconstruction without dense correspondences. *CoRR*, abs/1704.00337, 2017.

[45] Jian Shi, Yue Dong, Hao Su, and Stella X. Yu. Learning non-lambertian object intrinsics across shapenet categories. In *CVPR*, pages 5844–5853. IEEE Computer Society, 2017.

[46] Antonin Sulc, Anna Alperovich, Nico Marniok, and Bastian Goldluecke. Reflection separation in light fields based on sparse coding and specular flow. pages 137–144, 2016.

[47] Borom Tunwattanapong, Graham Fyffe, Paul Graham, Jay Busch, Xueming Yu, Abhijeet Ghosh, and Paul Debevec. Acquiring reflectance and shape from continuous spherical harmonic illumination. *ACM Trans. Graph.*, 32(4):109:1–109:12, July 2013. ISSN 0730-0301.

[48] V. Vaish, B. Wilburn, N. Joshi, and M. Levoy. Using plane + parallax for calibrating dense camera arrays. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 1, pages I–2–I–9 Vol.1, June 2004.

[49] Thomas Whelan, Renato F. Salas-Moreno, Ben Glocker, Andrew J. Davison, and Stefan Leutenegger. Elasticfusion: Real-time dense SLAM and light source estimation. *I. J. Robotics Res.*, 35(14):1697–1716, 2016.

[50] J. Wulff, D. J. Butler, G. B. Stanley, and M. J. Black. Lessons and insights from creating a synthetic optical flow benchmark. In A. Fusiello et al. (Eds.), editor, *ECCV Workshop on Unsolved Problems in Optical Flow and Stereo Estimation*, Part II, LNCS 7584, pages 168–177. Springer-Verlag, October 2012.

[51] Dehua Xie, Shuaicheng Liu, Kaimo Lin, Shuyuan Zhu, and Bing Zeng. Intrinsic decomposition for stereoscopic images. In *IEEE ICIP*, 2016.

[52] Genzhi Ye, Elena Garces, Yebin Liu, Qionghai Dai, and Diego Gutierrez. Intrinsic video and applications. *ACM Trans. Graph.*, 33(4), July 2014.

[53] Lap-Fai Yu, Sai Kit Yeung, Yu-Wing Tai, and Stephen Lin. Shading-based shape refinement of RGB-D images. In *CVPR*, pages 1415–1422. IEEE Computer Society, 2013.

[54] Qi Zhao, Ping Tan, Qiang Dai, Li Shen, Enhua Wu, and Stephen Lin. A closed-form solution to retinex with nonlocal texture constraints. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(7):1437–1444, 2012.

[55] Tinghui Zhou, Philipp Krähenbühl, and Alexei A Efros. Learning data-driven reflectance priors for intrinsic image decomposition. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3469–3477, 2015.

[56] M. Ziegler, R. op het Veld, J. Keinert, and F. Zilly. Acquisition system for dense lightfield of large scenes. In *2017 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, June 2017.