BRDF-Reconstruction in Photogrammetry Studio Setups

Matthias Innmann[1] Jochen Süßmuth[2] Marc Stamminger[1]

[1]Friedrich-Alexander University Erlangen-Nürnberg [2]adidas AG

Abstract

Photogrammetry Studios are a common setup to acquire high-quality 3D geometry from different kinds of real-world objects, humans, etc. In a photo studio like setup, 50 - 200 DSLR cameras are used with object-specific illumination to simultaneously capture images that are processed by algorithms that automatically estimate the camera parameters and detailed geometry. These steps are automated in established pipelines to a large extent and do not require much user input. However, the post-processing typically involves a manual estimation of surface reflectance parameters by an artist, who paints textures to allow for photorealistic rendering. While professional light stages facilitate this process in an automated way, these setups are very expensive and require accurately calibrated light sources and cameras. In our work, we present a new formulation along with a practical solution to reduce these constraints to photo studio like setups by jointly reconstructing the geometric configuration of the lights along with spatially varying surface reflectance properties and its diffuse albedo. In the presented synthetic as well as real-world experiments, we analyze the effect of different optimization objectives and show that our method is able to provide photorealistic reconstruction results with an RMSE of $\approx 1 - 3\%$ *on real data.*

1. Introduction

Today, photogrammetry studios are widely used to create 3D content for advertisement, engineering, or movie production. In such a studio, an array of DSLR cameras synchronously takes a picture of an object in the center. Structure-from-motion and multi-view stereo are used to automatically reconstruct the subject matter. While such setups achieve high-quality 3D geometry, much less attention has been paid to the reconstruction of the reflection properties, and mostly only a texture atlas is generated with baked diffuse lighting. To generate a texture with material parameters, a time-consuming and expensive manual material painting process follows. Alternatively, much more expensive light stage setups can be used, where multiple images of the object are taken with varying, controlled illumination.

In this work, we show how to extend the standard capturing process by an automatic reconstruction of a high-quality spatially varying BRDF (SVBRDF) of the subject matter. Our approach integrates easily into the typical workflow and requires only little extra effort: It is based on the RGB input images that are used for geometry reconstruction, in particular we do not need multiple images with varying illumination. Because lighting is usually adapted for each object to be captured, we additionally capture a single HDR environment map, roughly from the position of the object, which gives us all necessary information about the lighting setup.

Our process starts once the geometry of the object has been reconstructed using the normal workflow. In an optimization process we optimize spatially varying BRDF parameters based on the registered input RGB images. We determine parameters of the Disney BRDF [6], because it is widely used in production and can express different types of materials well. This optimization requires knowledge about the illumination of the object, which we take from the HDR panorama. However, the panorama image is only roughly aligned (for practical reasons), and light sources can be rather close to the object, so that the infinite illumination assumption does not hold. Thus, we optimize the alignment of the panorama, as well as the distance of the light sources visible in the panorama, jointly with the BRDF parameters. As a result, we are able to reconstruct spatially varying Disney BRDF parameters that reproduce the input images with only very little error, and that result in convincing renderings under arbitrary illumination and novel views.

In summary, we present the following contributions:

- A framework to estimate a spatially varying Disney BRDF from a set of images of an object under known, distant illumination from an environment map.
- An optimization approach to spatially align the environment map, and to reconstruct the position of the light sources.
- A pipeline for joint reconstruction of spatially varying albedo, reflectance and the geometric configuration of the light setup, which is able to provide photorealistic results on real-world data.
- Thorough evaluation of the proposed environment reconstruction method on a synthetic dataset as well as the influence of different optimization objectives.

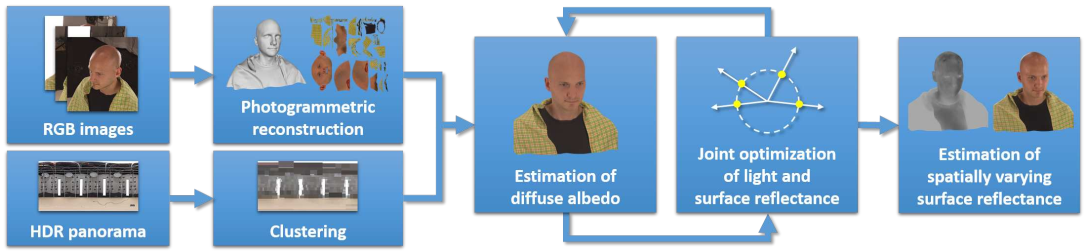


Figure 1: **Method overview:** From a set of RGB images, we run a photogrammetric reconstruction along with a clustered representation of a captured HDR panorama of the studio. Based on these data, we solve for the diffuse albedo, the light setup of the studio as well as the spatially varying surface reflectance.

2. Related Work

The reconstruction of geometry and material of real world objects is a long-standing and well researched problem, and often considered as the holy grail of visual computing. A good overview on related techniques is given in the state-of-the-art report by Guarnera et al. [12].

Related work can be categorized with respect to the structure of the input images and the acquisition setup.

RGB / RGB-D Video input Dong et al. [11] and Xia et al. [30] recover a spatially varying BRDF from a video of a rotating object of known geometry under unknown illumination. A general facet-based BRDF is determined, as well as the illumination. Xia et al. [30] additionally refine the shape to better explain the observations. Kim et al. [14] determine an SVBRDF with a KinectFusion pipeline: For each reconstructed voxel, observed color values are gathered, and a per-voxel BRDF is determined by a learning-based approach.

Controlled Illumination Lensch et al. [17] reconstruct SVBRDFs of an object with known geometry from a number of views with varying illumination from a point-like light source. Pose and light source are estimated from the images and a Lafortune BRDF is fitted [16]. In light stage setups [9, 29] multiple images from an object are made under varying, controlled illumination from a dome of light sources. These setups generate high-quality solutions, but the dome installation is expensive, and the capturing process is time-intensive. In other approaches, a simpler setup is used to generate controlled illumination, e.g. an LCD screen [4], a mini light stage [13], or a colored panel [27]. However, these setups can only handle flat material probes or small objects. Nielsen et al. [22] take a general look at the problem and present an approach to optimally select samples for BRDF acquisition in controlled environments.

Single (flash/no-flash) images A more recent stream of research take as input a single (or two) images, often with an off-the-shelf mobile phone with built-in flash to reconstruct an SVBRDF, yet they are limited to a flat light probe [5, 3, 18, 10], or to a single material [20]. Li et al. [19] reconstruct an SVBRDF from a single picture of non-flat objects, however naturally not the complete shape can be reconstructed but a depth buffer for the given view.

Unstructured Input Panagopoulus et al. [25] present a statistical method to estimate illumination directions and shadows from a single image. Zollhöfer et al. [31] refine geometry by estimating albedo and illumination. This method is extended by Kim et al. [15] also allowing for varying illumination between different images. While these methods achieve compelling results on a geometric level, both methods assume simple Lambertian reflectance and distant illumination, thus rendering them unsuitable for our objective. Ono et al. [23, 24] use Multi-View stereo to reconstruct the geometry and directly measure the BRDF (without any BRDF model) from a set of reliable samples. This method, however, is limited to homogeneous materials (single BRDF). Wang et al. [28] jointly estimate material and illumination from uncontrolled internet images, under the assumption that multiple images of the same object are available under varying illumination or varying view point. In a very recent work, Nam et al. [21] reconstruct geometry and SVBRDFs from a set of images, all taken with the built-in flash of a hand-held camera.

3. Approach

3.1. Setup and Input

Our work is based on a typical photogrammetry studio setup. An array of DSLR cameras, which does not have to be calibrated, takes about 50-100 high-quality images of an object. Typical subject matters are persons or other objects with no severe occlusions. Because cameras are synchronized, the scene does not have to be static, which makes it well applicable to persons. As in a photo studio, the object is illuminated by light panels, which can be rearranged individually for every new object. Based on these 50-100 input images, high-quality reconstructions are generated without user input, using multi-view stereo software like PhotoScan [2] or RealityCapture [26]. Most often, the obtained models are cleaned up manually, before they are fed to a standard content creation pipeline.

The generated models are of high quality, but usually include only textures with baked lighting. Thus, artists have to paint new textures with manually estimated material parameters, which is a lengthy and costly process, and often gives non-reproducible results. Goal of this paper is to automate this process, i.e., to automatically reconstruct high-quality SVBRDFs from such input. In some sense, our setup is comparable with a light stage, however, we do not need an expensive illumination dome, and can handle moving objects, because we use synchronous single shots from each camera. Due to the automatic registration of the cameras and the joint estimation of the light geometry, our setup is suitable for both object-dependent views and illumination.

Input to our method are:

- A high-quality reconstruction of the 3D geometry of the subject matter.
- The images of all cameras, including their precise extrinsic and intrinsic parameters.
- An HDR environment map of the setup, showing all light sources, taken roughly from the position of the subject matter.

We develop all the object images to 16 bit linear TIFF using dcraw [7]. The HDR environment map is captured using exposure bracketing. It is taken roughly from the position of the subject matter, its precise position and orientation are optimized during the reconstruction process.

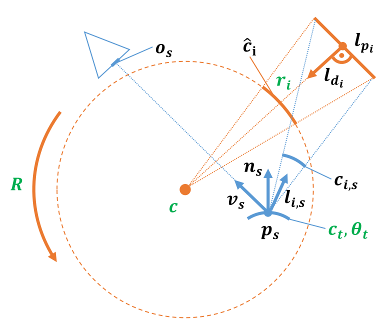
3.2. Representation of Illumination

Reconstructing the BRDFs makes it necessary to estimate illumination. In our experiments, it turned out that a photogrammetric reconstruction of the environment based on the input images does not provide useful results (see supplemental document). We thus use a single HDR environment map as basis, where each pixel can be seen as a single point light. To reduce the overall amount of variables, we create a low-dimensional representation, using the median cut approach by Debevec et al. [8], which allows us to approximate the environmental light using a smaller amount, e.g. 256, of directional light sources.

Because we don't know its precise position and orientation from which the environment map has been taken, we further add the center $\mathbf{c}$ and rotation $\mathbf{R}$ to the set of parameters to optimize.

Furthermore, we observe that the often made infinite distance assumption, as it is also typically done by methods that approximate incident lighting by Spherical Harmonics, e.g. [31, 15], does not hold for our environment maps: light panels are at a distance of 1m to 3m in our setups, which is close, relative to object size. We thus assign a distance parameter r_i to each light source i and also add these to the set of free variables.

Given an initial estimate of the direction of a light source $\hat{\mathbf{l}}_{d_i}$, the corresponding rotated light direction $\mathbf{l}_{d_i}$ and its position $\mathbf{l}_{p_i}$ are computed as:



s	sample index
t	texture index
i	light index
$\mathbf{p}_s$	position of s
$\mathbf{n}_s$	normal of s
$\mathbf{v}_s$	view direction of s
$\mathbf{o}_s$	observed color of s
$\mathbf{c}_t$	diffuse albedo at texel t
θ_t	BRDF at texel t
$\mathbf{c}$	center of env. map
$\mathbf{R}$	rotation of env. map
$\mathbf{l}_{d_i}$	env. map dir. of light i
$\hat{\mathbf{c}}_i$	env. map color of light i
$\mathbf{l}_{p_i}$	position of light i
$\mathbf{l}_{i,s}$	direction of light i at s
$\mathbf{c}_{i,s}$	color of light i at s

Figure 2: **Scene overview**: Our scene representation and the used notation throughout our paper (unknowns that we estimate are highlighted in green).

$$\mathbf{l}_{d_i} = \mathbf{R}\hat{\mathbf{l}}_{d_i}, \qquad \mathbf{l}_{p_i} = \mathbf{c} + r_i \mathbf{l}_{d_i}. \tag{1}$$

For a sample with position $\mathbf{p}_s$ and the observed color $\hat{\mathbf{c}}_i$ for light i in the environment map, the position dependent light direction $\mathbf{l}_{i,s}$ and its color $\mathbf{c}_{i,s}$ are:

$$\mathbf{l}_{i,s} = \frac{\mathbf{l}_{p_i} - \mathbf{p}_s}{\|\mathbf{l}_{p_i} - \mathbf{p}_s\|_2}, \tag{2}$$

$$\mathbf{c}_{i,s} = \frac{r_i^2}{\|\mathbf{l}_{p_i} - \mathbf{p}_s\|_2^2} \cdot \langle \mathbf{l}_{i,s}, \mathbf{l}_{d_i} \rangle_0 \cdot \hat{\mathbf{c}}_i \tag{3}$$

The term $\langle \mathbf{l}_{i,s}, \mathbf{l}_{d_i} \rangle_0$ accounts for the disk approximation of the light source, which offers a realistic representation.

For simpler notation, we define

$$\langle \mathbf{a}, \mathbf{b} \rangle_0 := \max\{\langle \mathbf{a}, \mathbf{b} \rangle, 0\}. \tag{4}$$

An overview of the scene representation and parametrization is given in Fig. 2.

3.3. Image Formation Model and BRDF Estimation

In the given setup with mostly convex objects, direct light is the dominant factor for surface shading. We therefore model the light transport considering only the incoming light of the light sources captured in the HDR panorama and ignore indirect illumination. To model surface reflectance, we use the BRDF described by Burley et al. [6]. This *Disney BRDF* is both physically based and highly expressive. Furthermore, it used in many real-world pipelines. The complete formulation of the Disney BRDF is given in the supplemental document.

For a sample s with its associated texel t, our image formation model for N lights is given by:

$$\sum_{i=1}^{N} \mathbf{f}(\mathbf{n}_s, \mathbf{v}_s, \mathbf{l}_{i,s}, \mathbf{c}_t | \theta_t) \cdot \mathbf{c}_{i,s}. \tag{5}$$

$\mathbf{f}$ is the used BRDF representation, in our case the Disney BRDF, with its parameters $\theta_t \in [0,1]^{10}$. The operator $(\cdot)$ refers to element-wise multiplication of the RGB values.

In our case, we aim to estimate the set $\mathcal{T}$ of surface reflectance parameters $\theta_t^\star$, its diffuse albedo $\mathbf{c}_t^\star$ and the geometric configuration of the light setup for N lights, i.e. the center $\mathbf{c}^\star$ and rotation $\mathbf{R}^\star$ of the captured HDR panorama and the distance $r_i^\star$ of each light source i to the center. We cast this as a non-linear optimization problem, which minimizes the dense photometric error w.r.t. to the observations $\mathbf{o}_s$ of the input RGB images:

$$E_{\text{photo}} = \sum_{t\in\mathcal{T}, s\in\mathcal{S}_t} \left\| \sum_{i=1}^{N} \mathbf{f}(\mathbf{n}_s, \mathbf{v}_s, \mathbf{l}_{i,s}, \mathbf{c}_t|\theta_t) \cdot \mathbf{c}_{i,s} - \mathbf{o}_s \right\|_2^2, \tag{6}$$

$$\{\theta_t^\star, \mathbf{c}_t^\star, \mathbf{c}^\star, \mathbf{R}^\star, r_i^\star\} = \operatorname*{arg\,min}_{\theta_t, \mathbf{c}_t, \mathbf{c}, \mathbf{R}, r_i} E_{\text{photo}}, \tag{7}$$

where $\mathcal{T}$ is the set of albedo/BRDF texels and $\mathcal{S}_t$ are the samples associated to texel t.

3.4. Joint Estimation of Albedo, Reflectance and Light

Initial Albedo Estimation Our algorithm is bootstrapped by first assuming directional lights, i.e., light sources that have a distance of infinity. The initial rotation $\mathbf{R}$ is obtained by a manual rough alignment of the environment map. We observe that the diffuse amount of the albedo is contained in the samples with the lowest luminosity.

Therefore, we compute the diffuse albedo $\mathbf{c}_t$ for all texels based on the quartile of observations $\mathbf{o}_s$ of the surface element having the lowest luminosity by solving the linear least-squares problem

$$\mathbf{c}_t^\star = \operatorname*{arg\,min}_{\mathbf{c}_t} \sum_{t\in\mathcal{T}, s\in\mathcal{S}_t} \left\| \sum_{i=1}^{N} \frac{1}{\pi} \langle \mathbf{n}_s, \hat{\mathbf{l}}_{d_i} \rangle_0 \cdot \mathbf{c}_t \cdot \hat{\mathbf{c}}_i - \mathbf{o}_s \right\|_2^2. \tag{8}$$

For this initial step, we deal with diffuse reflectivity and distant light sources and thus do not have to estimate any parameter for the reflectance (θ_t) or the lights ($\mathbf{R}, \mathbf{c}, r_i$).

Alternating Optimization There is an inherent ambiguity between light positions and albedo values: for a given set of light positions, there is a unique albedo texture and for a given texture, there is a unique set of light positions.

We observe that while the samples with the lowest luminosity represent the diffuse amount of the albedo, samples with high luminosity represent the specular highlights and can be used to triangulate the light sources. These specular samples are assumed to be of the same BRDF, i.e., we only use one parameter set θ.

We decouple our joint optimization problem of albedo, reflectance and light into two sub-problems with opposing properties and use an alternating optimization, which is iterated until convergence:

1. Given the current diffuse albedo, estimate the optimal reflectance ($\theta^\star$) and light setup ($\mathbf{R}^\star, \mathbf{c}^\star, r_i^\star$) from the high-luminosity samples:

$$\{\theta^\star, \mathbf{c}^\star, \mathbf{R}^\star, r_i^\star\} = \operatorname*{arg\,min}_{\theta, \mathbf{c}, \mathbf{R}, r_i} E_{\text{photo}}. \tag{9}$$

2. Given the current light, estimate the optimal diffuse albedo ($\mathbf{c}_t^\star$) from the low-luminosity samples assuming diffuse reflectivity:

$$\mathbf{c}_t^\star = \operatorname*{arg\,min}_{\mathbf{c}_t} E_{\text{photo}}, \tag{10}$$

with $\mathbf{f}(\mathbf{n}_s, \mathbf{v}_s, \mathbf{l}_{i,s}, \mathbf{c}_t|\theta) = \frac{1}{\pi}\langle \mathbf{n}_s, \mathbf{l}_{i,s} \rangle_0 \cdot \mathbf{c}_t$.

While problem 1 consists of only few variables ($N + 16$, with N being the number of light sources), its structure is very dense and highly non-linear. On the other hand, problem 2 consists of several millions of parameters (e.g. 12.6M for an albedo texture of 2048×2048), but it is extremely sparse and can be solved for each texel independently.

Similarity of Lights In our synthetic experiments (Sec. 4.1), we observe that even for ground-truth settings, the optimization of step 1 will result in incorrect light configurations as long as the reconstruction operates with a different number of light sources than the amount the data has been generated with, which is especially true for real data.

To mitigate this problem, we add a weighted similarity term to the optimizer that prefers configurations with similar distances (r_i, r_j) for neighboring light sources:

$$E_{\text{sim}} = \sum_{i=1}^{N} \sum_{j\in\mathcal{N}_i} \frac{1}{\|\hat{\mathbf{l}}_{d_i} - \hat{\mathbf{l}}_{d_j}\|_2^2} (r_i - r_j)^2. \tag{11}$$

This term enforces stronger similarity for light distances that have similar direction. Due to the preprocessing of the HDR panorama, the density of the clusters is proportional to the local luminosity of the environment map, i.e., the density resembles the actual lights. If light directions are closer to each other, the probability of being the same light source is high. On the other hand, if light directions are far apart, they are probably different light sources.

Given this additional constraint, we extend problem 1 to

$$\{\theta^\star, \mathbf{c}^\star, \mathbf{R}^\star, r_i^\star\} = \operatorname*{arg\,min}_{\theta, \mathbf{c}, \mathbf{R}, r_i} E_{\text{photo}} + w_{\text{sim}} E_{\text{sim}}. \tag{12}$$

3.5. Estimation of High-Quality Spatially Varying Reflectance Properties

We use the reconstructed light geometry (Sec. 3.4) to estimate high-quality spatially varying surface reflectance $\theta_t^\star$ (SVBRDFs) by minimizing the photometric energy. All pixels from every image that correspond to an observation of the object are used (overall up to 100M observations).

To reduce artifacts for undersampled surface areas and mitigate overfitting, we add a local smoothness regularizer in texture space to constrain the reflectance parameters:

$$E_{\text{reg}} = \sum_{t_i \in \mathcal{T}} \sum_{t_j \in \mathcal{N}_{t_i}} \|\theta_{t_i} - \theta_{t_j}\|_2^2, \tag{13}$$

where $\mathcal{N}_{t_i}$ are the neighboring reflectance parameters.

We solve the joint optimization problem

$$\theta_t^\star = \arg\min_{\theta_t} E_{\text{photo}} + w_{\text{reg}} E_{\text{reg}} \tag{14}$$

using a Levenberg-Marquardt (LM) solver in a hierarchical way (texture dimensions 1×1 to 2048×2048) to avoid getting stuck in local minima. For optimal precision and performance, we use analytic derivatives, wich we provide in the supplemental document.

Neighbors across texture boundaries As the surface parametrization output of regular photogrammetry software is a texture atlas, we use a surface-based approach to find neighbors for texels at boundaries. For every boundary point that has k neighbors in the texture, we find the $(4-k)$-nearest neighbors on the surface in any other texture segment and use the corresponding parameters as neighboring values for the regularization term.

3.6. Implementation

To build a system that achieves practical runtime, we use optimized implementations for two sub-problems: The linear-least squares problem to estimate the diffuse albedo (Eq. (10)), is solved entirely on the GPU.

Furthermore, we use a hand-crafted CUDA implementation utilizing shared memory and intrinsic functions, e.g. `__shfl_down`, to efficiently evaluate the sums in the entries of the Jacobian matrix and the residual vector. In our real-world datasets, we deal with up to 10^9 residuals and 40M variables – though $\mathbf{J}^T\mathbf{J}$ is extremely sparse, i.e. 14 non-zero entries per row. To solve the linear least-squares problem in each LM step, we use a parallel PCG solver on the CPU, which has negligible runtime (3.1%) compared to the evaluation of the Jacobian matrix and the residual vector.

A naive implementation, e.g., using the Ceres optimization framework [1] that utilizes automatic differentiation shows an increased runtime of $\approx 400\times$ in our experiments. While our GPU implementation runs for $\approx$ 3.5 hours, such implementation would result in a runtime of $\approx$ 60 days.

4. Evaluation

We conduct a thorough evaluation on both synthetic as well as real-world data. Furthermore, we analyze the influence of the regularization parameters in terms of accuracy as well as generalization of the reconstructed reflectance parameters. Our evaluation is done consistently in terms of photometric RMSE normalized to the RGB cube $[0, 1]^3$, i.e. $(0, 0, 0)^T$ vs. $(1, 1, 1)^T$ results in an error of 100%.

Several additional evaluations on both synthetic and real-world data can be found in the supplemental document.

4.1. Quantitative Evaluation on Synthetic Data

We create several studio-like datasets with varying complexity (see Fig. 3) using 85 images rendered from camera positions similar to the photogrammetry studio setup. Dataset 1 and 2 are created with diffuse shading. The images of dataset 3 are shaded using the Disney BRDF (Fig. 3e) with parameters $\theta = (0.2, 0.2, 0.5, 0.6, 0.9, 0.0, 0.3, 0.4, 0.2, 0.2)^T$ (described in more detail in the supplemental document).

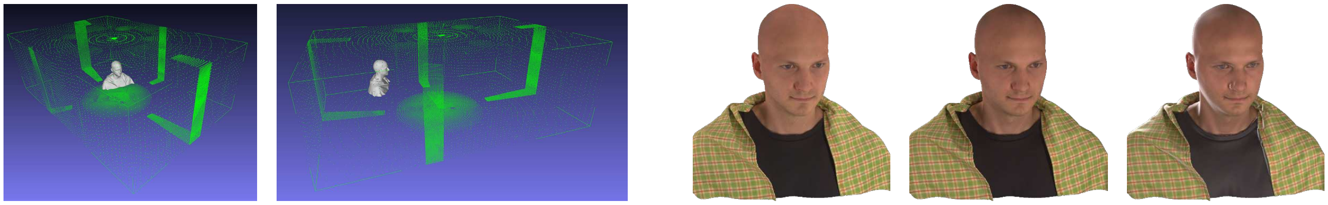
For dataset 1, the object is placed in center, whereas for dataset 2 and 3, the object is placed off-center.

Regularization of Light Sources We perform an extensive evaluation of the similarity objective E_{sim} (Sec. 3.4). To this end, we use dataset 1 (Fig. 3c) created with different amounts of light sources $\hat{N} \in \{256, 512, 65536\}$ on the cuboid ($3.6m \times 2.4m \times 1.2m$).

For each case, the reconstruction is performed with $N = 256$ light sources with a different value of w_{sim}. $\hat{N} = 256$ is the validation case, $\hat{N} = 512$ simulates an environment with similar number of light sources and $\hat{N} = 65536$ simulates a realistic environment with a high number of light sources. Furthermore, we initialize the distances with different values: 1m, the ground truth and 2m.

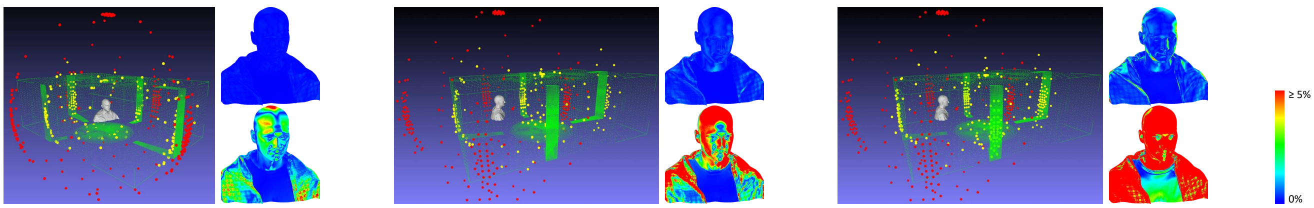
The results of this evaluation can be found in Fig. 5. The most important observation is two-fold: On the one hand, if we reconstruct the light geometry with the same amount of lights, i.e. 256, the optimization converges even for very small similarity weights for different types of initial values. On the other hand, if the number of light sources differs, e.g. 512 and 65536 for data creation, 256 for reconstruction, the optimization diverges even for ground truth initialization. This shows the importance of the additional similarity objective: As can be seen, a weight of $w_{\text{sim}} \in [10^{-7}, 10^{-6}]$ offers stable results for the light reconstruction, i.e., a relative distance error of $\approx 5\%$, with a low photometric RMSE.

Furthermore, we evaluate the joint reconstruction of center $\mathbf{c}$, distances r_i (dataset 2 and 3) and BRDF θ (dataset 3, all Fig. 3). The aim of this experiment is to evaluate if the optimization can reconstruct the center and the surface reflectance. Therefore, the object of interest, which typically



(a) Light config 1 (b) Light config 2 (c) Dataset 1 (d) Dataset 2 (e) Dataset 3

Figure 3: **Overview of synthetic datasets**: For two different light configurations (object in center and off-center; lights shown in green) we create three datasets. (1) is created with config 1 and diffuse reflectance, (2) is created with config 2 and diffuse reflectance, (3) is created with config 2 and a complex Disney BRDF.



(a) Dataset 1, RMSE: 0.08% // 1.8% (b) Dataset 2, RMSE: 0.22% // 4.6% (c) Dataset 3, RMSE: 0.67% // 11.6%

Figure 4: **Results for synthetic data**: We show the reconstructed light geometry (yellow) for $w_{\text{sim}} = 10^{-6}$ with its initialization (red) and the ground truth (green) along with the photometric error for one example image (upper). For baseline comparison, we also show the error using directional illumination without accurate reconstruction of the light sources (lower).

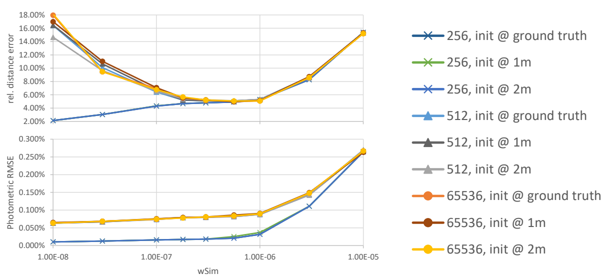


Figure 5: **Evaluation of light similarity term**: We evaluate the relative distance error of the light sources (upper) as well as the photometric RMSE (lower) for different values of w_{sim}, initializations and number of light sources $\hat{N}$.

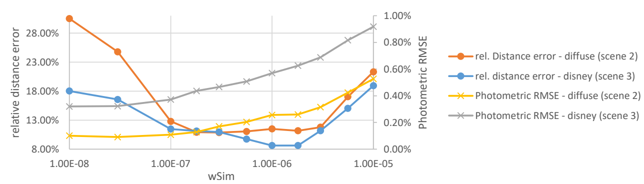


Figure 6: **Evaluation of light reconstruction**: We evaluate the accuracy of the reconstruction for different BRDFs. A complex Disney BRDF offers more shading cues to allow for a more precise reconstruction of the light sources.

serves as a rough estimate of the center of the environment map, is placed off-center. The input data is generated with 65536 light sources for a realistic simulation. The optimization is initialized with the center of the mesh's bounding box and distances of 2m, as large values guarantee better convergence (see supplemental document). For the reconstruction of dataset 3, all BRDF parameters are initialized to 0.5.

A visual overview of the reconstruction results can be found in Fig. 4, the quantitative evaluation is given in Fig. 6.

While the photometric RMSE for the data shaded with the Disney BRDF (Fig. 3e) is slightly higher than for the data with diffuse shading (Fig. 3d), the resulting light reconstruction is more precise using the Disney BRDF. I.e., for $w_{\text{sim}} = 10^{-6}$ the average error of the light positions is $\approx$ 10.7cm, with the center being off by only 0.79cm. For comparison, using the diffuse dataset, the average error is $\approx$ 13.9cm, with an error for the center of 1.45cm. This shows that more complex shading with dominating influence of distinct light sources and view-dependent shading leads to more accurate reconstruction of the light sources.

The reconstructed BRDF values are $\theta^\star = (0.195, 0.190, 0.493, 0.597, 0.892, 0.0, 0.314, 0.380, 0.202, 0.0)^T$. Apart from the last parameter (0.0 vs. 0.2), the estimated values have an average relative deviation of 2.1%, which shows that our optimization accurately recovers the surface reflectance. Regarding the last parameter (*clearcoatGloss*), which does not get reconstructed correctly, we make the following observations in our synthetic experiments: For values $\in [0, 0.8]$, it did not have any visible influence, only values $\in [0.8, 1]$ lead to visually different results. This observation can also be validated by inspecting the used GTR_1 distribution [6].

4.2. Evaluation on Real Data

To perform an evaluation on real-world data, we captured several datasets with different objects that are typical examples for photogrammetry studios (Fig. 7) and use RealityCapture [26] to reconstruct both the camera poses and the 3D geometry:

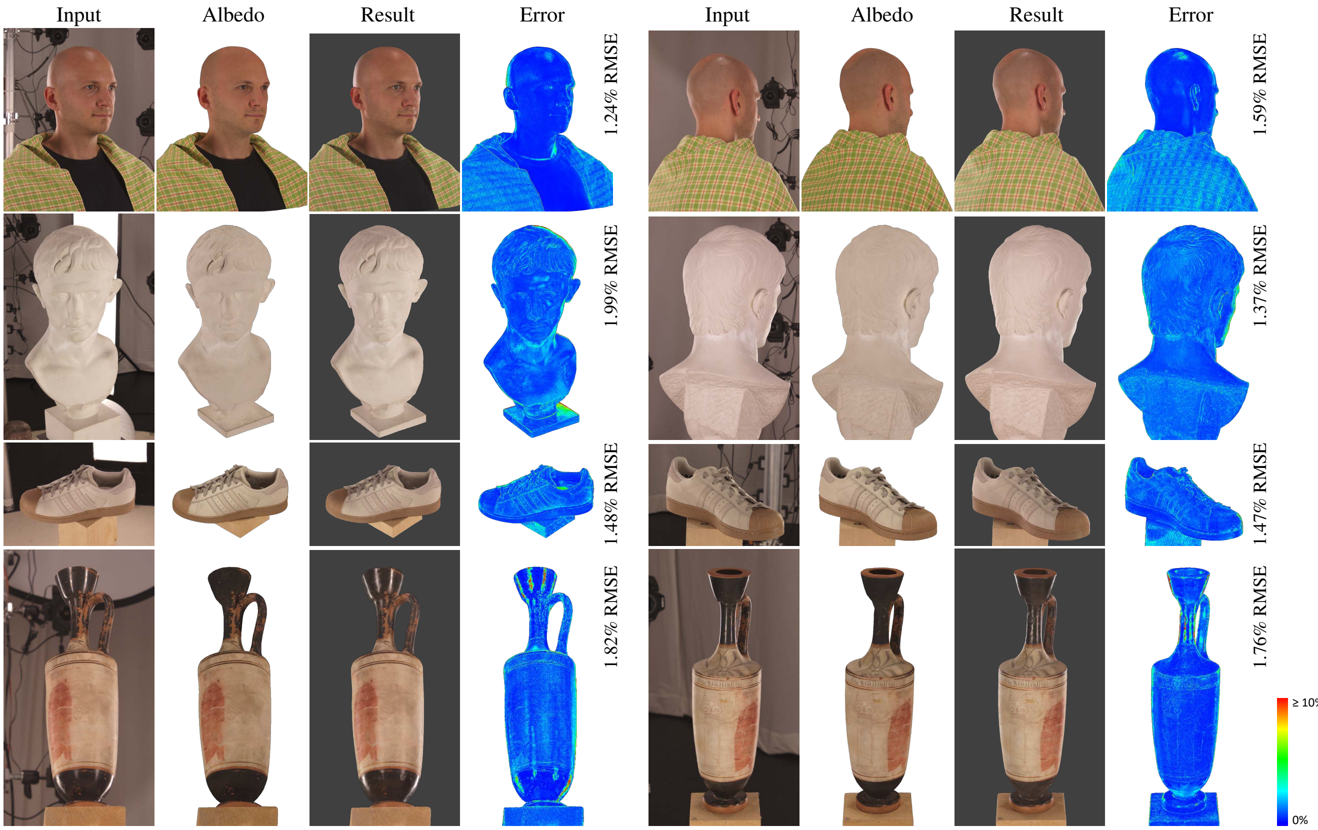


Figure 7: **Results on real-world data**: For the scenes UPPER BODY, BUST, SHOE and VASE, we show two example input images along with the estimated diffuse albedo and the re-rendered result as well as the photometric error.

1. A human UPPER BODY including a bald head, which serves as a challenging example considering the reflectance complexity of human skin [29].
2. An archaeological BUST that is of almost constant reflectance and albedo.
3. A SHOE that is a trade-off between complex geometry and complex surface reflectance.
4. An archaeological VASE that is of simpler geometry, but provides sharp reflectance and albedo boundaries as well as almost mirror-like surface areas.

Especially the UPPER BODY example is an interesting use-case for photogrammetry studios. While these setups are designed to output compelling geometry, photorealistic surface reflectance models often involve manual fine tuning of the parameters by an artist.

For several evaluations, we divide our input data into two disjoint sets of each 42 images. The set *fit* is used for the reconstruction of the parameters, *unseen* is used to evaluate the generalization of our method for unseen views. The joint error is evaluated on the complete input set *all*.

Using all input images, our method achieves an overall RMSE of $1.6\% - 2.3\%$ (Fig. 8). The most significant reduction of error occurs for the VASE scene due to the fact that it consists of distinct reflectance properties (almost dif-

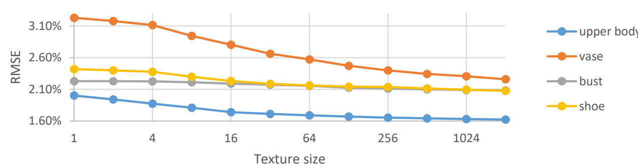


Figure 8: **Evaluation on real-world data**: We evaluate the overall error for different texture sizes.

fuse middle part and highly specular dark areas). The BUST shows almost no reduction as it is reasonable to assume to be of constant material with only slight differences over the surface (e.g. due to touching). While the scenes VASE, BUST and SHOE used the raw output of the photogrammetry reconstruction, a professional artist post-processed the geometry of the UPPER BODY example. This results in extremely accurate surface normals, which allows for superior performance in the SVBRDF estimation part (Sec. 3.5).

In Fig. 7, we show two example images (out of 85 total input images), the diffuse albedo and the resulting synthetically rendered image along with the photometric error. More renderings can be found in the accompanying video.

Texture boundaries The influence of our nearest neighbor search across textures is shown in Fig. 9. If the regularizer only operates inside each part of texture separately, the reconstructed SVBRDF exhibits discontinuities. This leads to shading artifacts for novel views. Using the regularization, these artifacts get mitigated, while the resulting photometric RMSE shows similar value (1.612% vs. 1.615%).

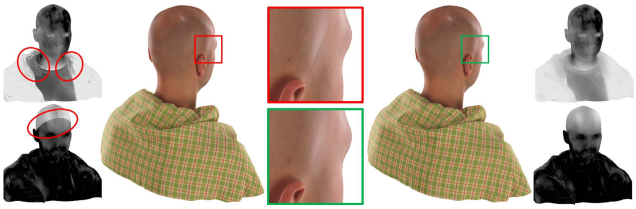


Figure 9: **Texture boundaries**: Without regularization across texture boundaries (left), the reconstructed SVBRDF exhibits discontinuities (two example layers shown), which cause shading artifacts for novel views (middle).

Comparison against Directional Lights We compare our method against the naive approximation of directional lights, i.e. lights of infinite distance, for the VASE dataset, which offers sharp reflectance boundaries and defined highlights. As can be seen in Fig. 10, our reconstruction of light positions (*point lights*) provides reduced photometric error compared to *directional lights* for all three sets (*fit, unseen, all*), given sufficient texture resolution, i.e. $\geq 64 \times 64$.

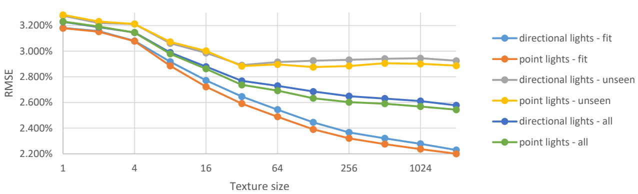


Figure 10: **Comparison against directional lights**: Using our method to reconstruct the geometric configuration of the lights (*point lights*), we are able to reduce the photometric error compared to a naive assumption of *directional lights*.

Regularization of Reflectance Parameters To evaluate the influence of our smoothness constraint for SVBRDFs (Sec. 3.5), we run our algorithm on the head of the UPPER BODY dataset for different values of w_{reg} (see Fig. 11)

A regularization weight $w_{\text{reg}} \approx 1 - 20$ offers similar results on unseen data, while $w_{\text{reg}} \approx 1$ performs best overall. For lower weights, the optimizer overfits to the seen data, whereas for higher weights, the property of spatially varying parameters gets lost and the result converges to a single set of parameters for $w_{\text{reg}} \to \infty$.

Using these BRDF textures layers, it is possible to synthesize photorealistic renderings of the object from novel views and under different illumination. We show several animations in the accompanying supplemental video.

5. Limitations and Future Work

While we are able to provide photorealistic results, it is important to point out limitations of our method and discuss how these problems can be addressed in future work.

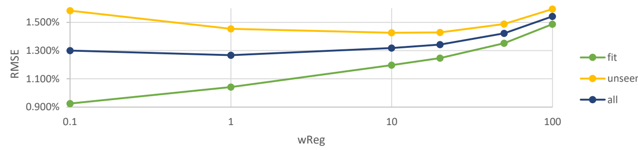


Figure 11: **Evaluation of BRDF regularizer**: We run the algorithm with varying strength w_{reg} of the smoothness constraint (defined in Sec. 3.5) and evaluate the photometric RMSE for the three different image sets.

Incorrect geometry causes severe errors especially along silhouettes (BUST) as well as in very glossy areas (VASE, both Fig. 7). We believe that our method to estimate the surface reflectance and the light setup can assist stereo-based 3D reconstruction to mitigate problems in these glossy areas and improve the accuracy of the surface normals. Furthermore, slightly incorrect camera alignment also leads to severe errors especially for areas with a high-frequency albedo, e.g., the colored cloth in the UPPER BODY example.

While the used Disney BRDF representation allows to model complex BRDFs, such as human skin, our method is currently not able to reconstruct arbitrary anisotropic materials. Extending the optimization and jointly solving for the anisotropy directions will thus allow for even more complex shading behavior.

Inherent to our image formation model, (self-)shadowing as well as indirect illumination is currently neglected. Therefore, these information get integrated into the albedo, which will cause slight artifacts for illumination changes, i.e., shadows will not be at the correct position anymore. Jointly solving these problems is challenging, but we believe that recent advancements especially in the domain of deep-learning can help to make these challenges tractable.

6. Conclusion

We presented a method to estimate albedo and spatially varying Disney BRDF parameters of an object in a photogrammetry setup. Illumination is reconstructed from a panoramic HDR image, where the position of the light sources is jointly optimized with the albedo and BRDF parameters. The approach generates high-quality Disney BRDF textures, that make high-quality rerenderings from arbitrary view points and under novel illumination possible. Our method thus can replace the time intensive and costly manual BRDF painting step, that is still necessary in current setups before the model can be used in production.

Acknowledgments

We thank Falko Matern for serving as a model for the UPPER BODY scene and Frank Zwick (optimission.de) for capturing and processing the datasets. This research is funded by the Bayerische Forschungsstiftung (For3D).

References

[1] S. Agarwal, K. Mierle, and Others. Ceres solver. http://ceres-solver.org. 5

[2] Agisoft. PhotoScan: MVS software, 2000–2004. 2

[3] M. Aittala, T. Aila, and J. Lehtinen. Reflectance modeling by neural texture synthesis. *ACM Transactions on Graphics (TOG)*, 35(4):65, 2016. 2

[4] M. Aittala, T. Weyrich, and J. Lehtinen. Practical svbrdf capture in the frequency domain. *ACM Trans. Graph.*, 32(4):110–1, 2013. 2

[5] M. Aittala, T. Weyrich, J. Lehtinen, et al. Two-shot svbrdf capture for stationary materials. *ACM Trans. Graph.*, 34(4):110–1, 2015. 2

[6] B. Burley and W. D. A. Studios. Physically-based shading at disney. In *ACM SIGGRAPH*, volume 2012, pages 1–7, 2012. 1, 3, 6

[7] D. Coffin. dcraw, 1997–2018. 3

[8] P. Debevec. A median cut algorithm for light probe sampling. In *ACM Siggraph 2005 Posters*, page 66. ACM, 2005. 3

[9] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar. Acquiring the Reflectance Field of a Human Face. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, pages 145–156. ACM Press/Addison-Wesley Publishing Co., 2000. 2

[10] V. Deschaintre, M. Aittala, F. Durand, G. Drettakis, and A. Bousseau. Single-image svbrdf capture with a rendering-aware deep network. *ACM Transactions on Graphics (TOG)*, 37(4):128, 2018. 2

[11] Y. Dong, G. Chen, P. Peers, J. Zhang, and X. Tong. Appearance-from-motion: Recovering spatially varying surface reflectance under unknown lighting. *ACM Transactions on Graphics (TOG)*, 33(6):193, 2014. 2

[12] D. Guarnera, G. C. Guarnera, A. Ghosh, C. Denk, and M. Glencross. Brdf representation and acquisition. In *Computer Graphics Forum*, volume 35, pages 625–650. Wiley Online Library, 2016. 2

[13] K. Kang, Z. Chen, J. Wang, K. Zhou, and H. Wu. Efficient reflectance capture using an autoencoder. *ACM Trans. Graph*, 37(4), 2018. 2

[14] K. Kim, J. Gu, S. Tyree, P. Molchanov, M. Nießner, and J. Kautz. A lightweight approach for on-the-fly reflectance estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 20–28, 2017. 2

[15] K. Kim, A. Torii, and M. Okutomi. Multi-view inverse rendering under arbitrary illumination and albedo. In *European Conference on Computer Vision*, pages 750–767. Springer, 2016. 2, 3

[16] E. P. F. Lafortune, S.-C. Foo, K. E. Torrance, and D. P. Greenberg. Non-linear Approximation of Reflectance Functions. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, pages 117–126. ACM Press/Addison-Wesley Publishing Co., 1997. 2

[17] H. P. A. Lensch, J. Kautz, M. Goesele, W. Heidrich, and H.-P. Seidel. Image-based Reconstruction of Spatial Appearance and Geometric Detail. *ACM Trans. Graph.*, 22(2):234–257, 2003. 2

[18] Z. Li, K. Sunkavalli, and M. Chandraker. Materials for masses: Svbrdf acquisition with a single mobile phone image. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 72–87, 2018. 2

[19] Z. Li, Z. Xu, R. Ramamoorthi, K. Sunkavalli, and M. Chandraker. Learning to reconstruct shape and spatially-varying reflectance from a single image. In *SIGGRAPH Asia 2018 Technical Papers*, page 269. ACM, 2018. 2

[20] A. Meka, M. Maximov, M. Zollhoefer, A. Chatterjee, H.-P. Seidel, C. Richardt, and C. Theobalt. Lime: Live intrinsic material estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6315–6324, 2018. 2

[21] G. Nam, J. H. Lee, D. Gutierrez, and M. H. Kim. Practical svbrdf acquisition of 3d objects with unstructured flash photography. In *SIGGRAPH Asia 2018 Technical Papers*, page 267. ACM, 2018. 2

[22] J. B. Nielsen, H. W. Jensen, and R. Ramamoorthi. On Optimal, Minimal BRDF Sampling for Reflectance Acquisition. *ACM Trans. Graph.*, 34(6):186:1–186:11, 2015. 2

[23] T. Ono, H. Kubo, T. Funatomi, and Y. Mukaigawa. Brdf reconstruction from real object using reconstructed geometry of multi-view images. In *SIGGRAPH Asia 2017 Technical Briefs*, page 8. ACM, 2017. 2

[24] T. Ono, H. Kubo, K. Tanaka, T. Funatomi, and Y. Mukaigawa. Practical brdf reconstruction using reliable geometric regions from multi-view stereo. *Computational Visual Media*, pages 1–12. 2

[25] A. Panagopoulos, D. Samaras, and N. Paragios. Robust shadow and illumination estimation using a mixture model. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 651–658. IEEE, 2009. 2

[26] C. Reality. Realitycapture, 2016–2019. 2, 6

[27] R. Vidaurre, D. Casas, E. Garces, and J. Lopez-Moreno. Brdf estimation of complex materials with nested learning. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1347–1356. IEEE, 2019. 2

[28] T. Wang, T. Ritschel, and N. Mitra. Joint material and illumination estimation from photo sets in the wild. In *2018 International Conference on 3D Vision (3DV)*, pages 22–31. IEEE, 2018. 2

[29] T. Weyrich, W. Matusik, H. Pfister, B. Bickel, C. Donner, C. Tu, J. McAndless, J. Lee, A. Ngan, H. W. Jensen, et al. Analysis of human faces using a measurement-based skin reflectance model. In *ACM Transactions on Graphics (TOG)*, volume 25, pages 1013–1024. ACM, 2006. 2, 7

[30] R. Xia, Y. Dong, P. Peers, and X. Tong. Recovering shape and spatially-varying surface reflectance under unknown illumination. *ACM Transactions on Graphics (TOG)*, 35(6):187, 2016. 2

[31] M. Zollhöfer, A. Dai, M. Innmann, C. Wu, M. Stamminger, C. Theobalt, and M. Nießner. Shading-based refinement on volumetric signed distance functions. *ACM Transactions on Graphics (TOG)*, 34(4):96, 2015. 2, 3