# Practical SVBRDF Acquisition of 3D Objects with Unconstructed Flash Photography

Finn Matras

Faculty of Informatics - Technische Universität München

**Abstract**

[12] proposes a practical method of acquiring both 3D geometry, SVBRDF and shading normals of an object using no specialized hardware. Only a camera with an active light source (flash) is needed as the images are unconstructed. The method is presented as a joint iterative reconstruction and inverse-rendering pipeline. Furthermore, the authors do not make any assumptions of only diffuse reflections. Thanks to this, input noise is suppressed and the resulting reconstructions contain high frequency details. Since all the stages of the process are intertwined and can leverage information from one another, the method is robust against low input image counts and converges nicely. In the conducted experiments the proposed method outperforms several existing state-of-the-art techniques, while also being more practical.

## 1 Introduction

An important part of computer graphics, be it for personal or professional use, is acquiring and reproducing the appearance of real-world objects. Many different methods for achieving this goal have been proposed. A common denominator for these methods is that there is a trade-off between performance and complexity (cost). While professional systems like the OTOY LightStage start at about 200.000$ [12], consumer products like the Shining 3D EinScan Pro 2X Plus cost less than 10.000$.

While most similar methods today require precomputed 3D models, complicated light-rigs or other advanced hardware, the method presented by [12] does not suffer from these limitations. It uses multiple unstructured images from a camera with flash. Key to the method is the joint reconstruction of the spatially varying bidirectional reflection distribution function (SVBRDF), shading normals and 3D geometry. A SVBRDF is a function that describes how light is reflected at different points on a surface. The algorithm is defined as an iterative and alternating optimization problem using inverse-rendering. A big advantage of the joint reconstruction is that one does not have to assume that the object only dissipates diffuse reflections, resulting in more high-frequency details in the final reconstruction.

## 2 Related work

Previous work done in this field mostly focus on acquiring and reproducing either appearance or geometry. Acquiring both at the same time usually requires special hardware like a gantry, light stage or a 3D scanner. Related work can be summarized in the following four categories:

- Reflectance capture of known 3D geometry:

  Methods using basis reflectancy functions and linear blending dependant on known 3D geometry. Such methods were presented by [8] and [17]. Both these methods require a commercial 3D scanner to accurately acquire the geometry of the 3D object.

- Reflectance capture of 2D planar geometry:

  Several methods requiring no specialized hardware have been proposed among others by [1] who use a camera and an LCD screen or [14] and [5] who proposed a method for reconstructing the appearance of near-flat objects with a smartphone camera. Methods requiring more complicated setups using for instance a computer-controlled illuminated LED-dome or a four axis spherical gantry were presented by [11] and [7] respectively.

- 3D reconstruction assuming diffuse reflection only:

  Using shading normals from the surface reflection one can enhance the geometric details of the 3D reconstruction. [4] and [13] proposed a method using Structure from Motion (SfM) and Multi-View Stereo (MVS) to update the geometry using estimated surface normals. Others propose similar methods to the above, but instead of SfM they use shape from silhuette in [3] and a signed distance function, SDF, in [18].

- Simultaneous acquisition of reflectance and 3D geometry:

  [2] use a light stage and discrete spherical illumination with polarized light sources. Other setups use a LED arm that rapidly orbits the object to create spherical illumination with harmonic patterns as presented by [15]. Many other methods exist, often requiring costly specialized hardware setup, an additional source of information given by for example multiple lights or spherical illumination. A large drawback is that acquisition often takes several hours.

# 3 Method description

The method presented in this report falls in the last category presented in the section above. In contrast to the methods discussed there, the presented method does not require any specialized hardware, a smartphone with a camera with flash is enough. As shown in the results, if the camera flash is bright enough, no special care has to be taken with regard to the ambient light. This makes the presented method easy to apply and separates it from previously proposed methods.

## 3.1 Image and reflectance model

Input to the algorithm are K unstructured flash photographs, $\mathbf{I} = \{I_k\}$. The light flash gives active illumination from which normals and reflectance are estimated. Each image is converted into linear radiance using Equation (1)

$$I(\mathbf{u}) = L(\mathbf{o}; \mathbf{x})\Delta t \Delta g \tag{1}$$

where $I(\mathbf{u})$ is the image at pixel $\mathbf{u}$, $L(\mathbf{o}; \mathbf{x})$ is the outgoing radiance from the 3D geometry point $\mathbf{x}$ in view direction $\mathbf{o}$, $\Delta t$ is the exposure time and $\Delta g$ is the flash intensity. The outgoing radiance is defined as shown in Equation (2)

$$L(\mathbf{o}; \mathbf{x}) = f(\mathbf{i}, \mathbf{o}; \mathbf{x}, \mathbf{n})L(-\mathbf{i}; \mathbf{x})(\mathbf{n} \cdot \mathbf{i}) \tag{2}$$

where $f(\mathbf{i}, \mathbf{o}; \mathbf{x}, \mathbf{n})$ is the SVBRDF function at point $\mathbf{x}$ with geometric normal $\mathbf{n}$, light direction $\mathbf{i}$ and view direction $\mathbf{o}$. $L(-\mathbf{i}; \mathbf{x})$ is the incident light at point $\mathbf{x}$. $(\mathbf{n} \cdot \mathbf{i})$ is a term that scales the reflectance so that it goes to zero as the light becomes perpendicular to the surface normal. The SVBRDF is defined as a set of $B$ basis BRDFs, $\mathbf{F_b} = \{f_b\}$ scaled by a weight map $\mathbf{W} = \{\omega_{p,b}\}$ to make then spacially variable. This is shown in Equation (3)

$$\mathbf{F} = \big\{f(\mathbf{i}, \mathbf{o}; \mathbf{x}, \mathbf{n})\big\} = \left\{ \sum_{b=1}^{B} \omega_{p,b} f_b(\mathbf{i}, \mathbf{o}) \right\} \tag{3}$$

The Cook-Torrance model is used for the basis BRDFs and is shown in Equation (4).

$$f_b(\mathbf{i}, \mathbf{o}) = \frac{\rho_d}{\pi} + \rho_s \frac{D(\mathbf{h})G(\mathbf{n}, \mathbf{i}, \mathbf{o})F(\mathbf{h}, \mathbf{i})}{4(\mathbf{n} \cdot \mathbf{i})(\mathbf{n} \cdot \mathbf{o})} \tag{4}$$

The first term describes the diffuse reflectance parametrized by the diffuse albedo $\rho_s$, while the specular albedo, $\rho_d$, scales the specular reflectance in the second term. The $D(\mathbf{h})$ is the distribution function, which describes how the reflectance is distributed on the geometry, the geometric term, $G(\mathbf{n}, \mathbf{i}, \mathbf{o})$, describes the surface of the object and how its texture affect reflected radiance. Finally the Fresnel function, $F(\mathbf{h}, \mathbf{i})$, describes how the light interacts with the surface at different angles.

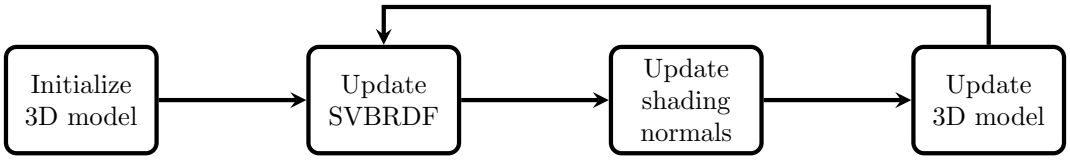


Figure 1: General overview of algorithm flow

## 3.2 Algorithm

A general overview of the proposed algorithm is shown in Figure 1. It starts with initialization of the 3D model. Then an estimate of the SVBRDF is computed, the shading normals are updated and finally the 3D geometry is updated. This process is repeated several times until the error converges. Due to the optimization, the images are separated into training and test data to avoid overfitting in a random 9:1 ratio.

### 3.2.1 Initialization

As camera parameters vary, one has to estimate several extrinsic camera parameters, among others the position of the camera-lens with respect to the flash light. This is done using images of chrome balls and a checkerboard. As these parameters do not change, this step has to be done only once for every device. The process is described by [9].

The first actual step of the algorithm is finding a base-mesh of the 3D geometry. First SfM is used to find the poses of all images, and then MVS is used to generate a pointcloud. Using screened Poisson surface reconstruction, [6], shown in Equation (5). The pointcloud is transformed into a low resolution voxel grid ($2^7$) to remove noise. The voxel grid is then given a higher resolution ($2^{10}$) which is matched to the actual 3D geometry using the estimated shading normals in later optimization steps.

$$\min_x \int \left|\left|\mathbf{V}(\mathbf{x}_p) - \nabla\chi(\mathbf{x}_p)\right|\right|^2 d\mathbf{x}_p + \alpha \sum_{\mathbf{x}_p \in \mathbf{X}} \chi(\mathbf{x}_p)^2 \tag{5}$$

### 3.2.2 Updating the SVBRDF

Updating the SVBRDF is a two step process. As previously defined, the SVBRDF consists of a set of basis BRDFs and a weight map. First, the basis BRDFs, $\mathbf{F_b}$, are reconstructed using the objective function shown in Equation (6). The NAG e04nkc [10] solver is used for this sparse quadratic programming problem.

$$\min_{\mathbf{F_b}} \sum_{p=1}^{P} \sum_{k=1}^{K} v_{p,k} \left( f'_{p,k} - \mathbf{\Phi}_{p,k}^T \sum_{b=1}^{B} \omega_{p,b} \mathbf{f_b} \right)^2 \tag{6}$$

$v_{p,k}$ is a binary visibility-weight, $\mathbf{\Phi}$ is a measurement vector consisting of scaled image samples and $f'$ is the captured reflectance. To get the new SVBRDF we still have to update the weight map, $\mathbf{W}$. This is done by minimizing the function as shown in Equation (7).

$$\min_{\omega_p} \frac{1}{2} \left|\left|\mathbf{Q}\boldsymbol{\omega_p} - \mathbf{r}\right|\right|^2 \quad s.t. \;\; \omega_{p,b} > 0, \;\; \sum_{b=1}^{B} \omega_{p,b} = 1 \tag{7}$$

where $\mathbf{Q}$ is a matrix of all $\mathbf{F_b}$s for all images, $\mathbf{r}$ is the observed reflectance and the constraints ensure unit scaling. To solve it, the convex quadratic programming solver NAG e04ncc [10] is used.

### 3.2.3 Updating the shading normals

To estimate the new shading normals, the geometric normals from the current 3D geometry are used as the initial estimate for the shading normals. Then Equation (8) is minimized to get the new estimates for the shading normals.

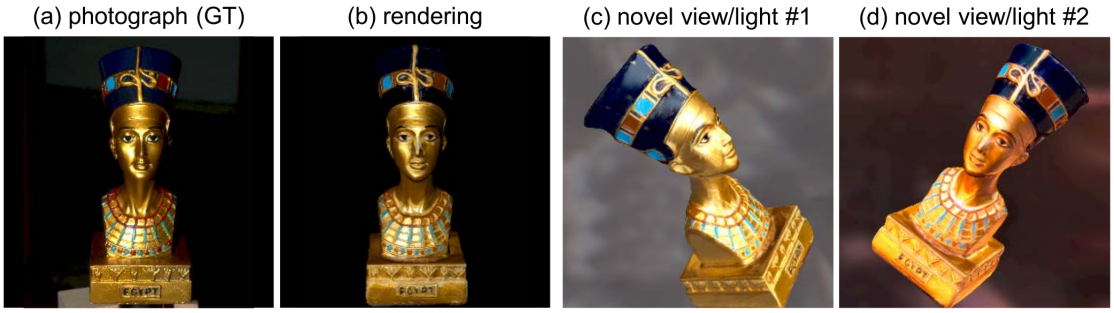


Figure 2: Acquisition and reconstruction of Nefertiti sculpture. Source: [12]

$$\min_{\mathbf{n}_p} \sum_{p=1}^{P} \sum_{k=1}^{K} v_{p,k} \left( L(\mathbf{o}_{p,k}; \mathbf{x}_p) - f(\mathbf{i}_{p,k}, \mathbf{o}_{p,k}; \mathbf{x}_p, \mathbf{n}_p) L(-\mathbf{i}_{p,k}; \mathbf{x}_p) \mathbf{n}_p \cdot \mathbf{i}_{p,k}) \right)^2 \tag{8}$$

#### 3.2.4 Updating the 3D geometry

Using the updated shading normals as the vector field $\mathbf{V}(\mathbf{x}_p)$ in the screened Poisson surface reconstruction described in Equation (5), a new 3D model geometry is computed.

## 4 Experiments and results

All the following experiments used either a Nikon D7000 DSLR camera or a LG Google Nexus 5X smartphone to capture the images, and a PC with an Intel i7-3770 3.40 GHz CPU with 32 GB of RAM and a NVIDIA GTX1080 GPU. In terms of runtime, the algorithm uses about 3-5 hours, where the initialization takes up most of the time. About 100 to 400 images covering the whole object are required for good results. The method was tested on a wide range of geometries and materials. An example is shown in Figure 2.

### 4.1 Influence of ambient light

As the method assumes that the only light reflected by the object comes from the active light source, naturally one would like to minimize the ambient light when capturing the input images. This is indeed the case when one does not have a strong flash, like in a smartphone. The DSLR on the contrary, has a very strong flash and ambient light has no significant influence on the resulting reconstruction.

### 4.2 Geometric accuracy

As stated previously, the presented method is geometrically accurate. In this experiment, a commercial desktop 3D scanner, the NextEngine (3000$), was used as reference, and the proposed method was compared to the COLMAP algorithm. The smartphone was used to capture the input images. As the COLMAP accuracy depends on the resolution, both high ($2^{10}$) and low ($2^7$) resolutions were tried. Figure 3 shows the geometric normals of the 3D reconstruction. As one can see from the close-ups, the proposed method has less noise than the COLMAP with high resolution, but a lot more details that the COLMAP with low resolution. It also reconstructs finer details than the reference model from the 3D scanner. The proposed method is about twice as good as COLMAP with the same resolution.

### 4.3 Impact of initial geometry

By comparing the final results to the initial geometries given by both the visual hull and COLMAP reconstructions, it could be seen that the visual hull reconstruction contains some noise, while

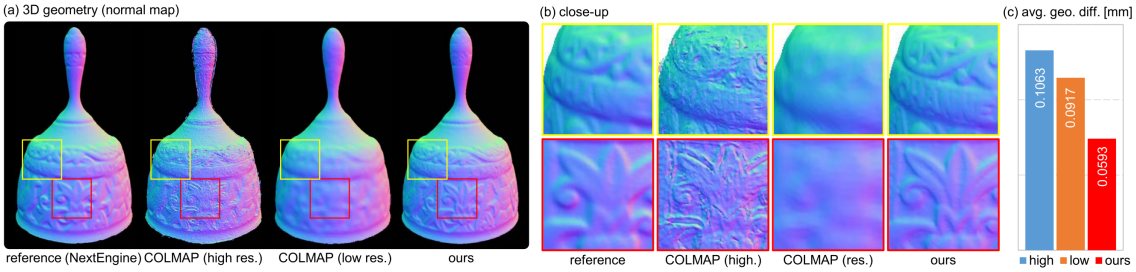


Figure 3: Comparison of geometric accuracy between a desktop 3D scanner, the COLMAP algorithm for two resolutions and the proposed method in [12]. Source: [12]

COLMAP clearly has less noise. As the proposed algorithm refines the initial geometry using the SVBRDF data to remove noise and inaccuracies from the initial 3D geometry, it delivers good and comparable results in both cases.

### 4.4 Impact of the number of basis BRDFs

A basis BRDF count of 9 was set based on empirical data, as this gives the best trade-off between computational complexity and reconstruction error.

### 4.5 Impact of the number of input images

The iterative optimization process works well with low image counts. However, as SfM fails at about 40 images, this is a lower limit for the whole method.

### 4.6 Still images vs. video frames

There exist related methods that use video frames instead of images. As active lighting is required, only the smartphone is able to capture video frames that satisfy this requirement. Due to the motionblur, inaccurate focus and lower dynamic range in the video frames, experiments showed that still images give noticeably better results.

## 5 Discussion of results

As the images are unstructured, it is not guaranteed that the whole surface will be captured well enough to represent all specular highlights. Other approaches presented by [5], [14] and [16] estimate BRDFs for each point on the 3D geometry independently. This requires at least one specular and one diffuse sample for each point. Using a set of basis BRDFs and weight maps as in this method mitigates this issue.

Even though the presented method works well in many scenarios, it still inherits some fundamental limitations of image-based 3D modeling techniques. This means that objects like pinecones, glass bottles and very shiny objects would not be reconstructed properly, as complex geometries, interreflections, subsurface scattering and transparency are not taken into account.

## 6 Summary

This is a novel solution to simultaneously reconstructing spatially-varying reflectance and 3D geometry using an off-the-shelf camera. The method gives good results by jointly formulating three reconstruction problems: SVBRDF, shading normals and 3D geometry. Other work usually requires more complex and expensive hardware, while the proposed method works better than some state-of-the-art methods for computing only one of the three reconstruction problems.

# References

[1] Miika Aittala, Tim Weyrich, and Jaakko Lehtinen. Practical svbrdf capture in the frequency domain. In *ACM Transactions on Graphics*, Vol. 32, No. 4, Article 110, 2013.

[2] Abhijeet Ghosh, Tongbo Chen, Pieter Peers, Syrus A. Wilson, and Paul Debevec. Circularly polarized spherical illumination reflectometry. In *ACM Transactions on Graphics (TOG)*, Dec. 2010.

[3] Carlos Hernandez, George Vogiatzis, and Roberto Cipolla. Multiview photometric stereo. In *IEEE Transactions on Pattern Analysis and Machine Intelligence 30, 3 (2008)*, page 548 – 554, 2008.

[4] Tomoaki Higo, Yasuyuki Matsushita, Neel Joshi, and Katsushi Ikeuchi. A hand- held photometric stereo camera for 3-d modeling. In *Computer Vision, 2009 IEEE 12th International Conference on. IEEE*, page 1234 – 1241, 2009.

[5] Zhuo Hui, Kalyan Sunkavalli, Joon-Young Lee, Sunil Hadap, Jian Wang, and Aswin C. Sankaranarayanan. Reflectance capture using univariate sampling of brdfs. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, page 5362 – 5370, 2017.

[6] Michael Kazhan and Hugues Hoppe. Screened poisson surface reconstruction. In *ACM Transactions on Graphics (TOG)*, vol. 32, No. 3, Article 29, 2013.

[7] Jason Lawrence, Aner Ben-Artzi, Christopher DeCoro, Wojciech Matusik, Hanspeter Pfister, Ravi Ramamoorthi, and Szymon Rusinkiewicz. Inverse shade trees for non-parametric material representation and editing. In *ACM Transactions on Graphics (TOG)*, Vol 25, No. 3, page 735 – 745, 2006.

[8] Hendrik PA Lensch, Jan Kautz, Michael Goesele, Wolfgang Heidrich, and Hans-Peter Seidel. Image-based reconstruction of spatially varying materials. In *Rendering techniques*, pages 103–114, 2001.

[9] Hendrik PA Lensch, Jan Kautz, Michael Goesele, Wolfgang Heidrich, and Hans-Peter Seidel. Image-based reconstruction of spatial appearance and geometric detail. In *ACM Transactions on Graphics (TOG)*, vol. 22, No. 2, pages 234–257, 2003.

[10] Koki Nagano, Graham Fyffe, Oleg Alexander, Jernej Barbic, Hao Li, Abhijeet Ghosh, and Paul E. Debevec. Numerical algorithms group: Skin microstructure deformation with displacement map convolution. In *ACM Transactions on Graphics (TOG)*, Vol. 34, No. 4, Article 109, 2015.

[11] Giljoo Nam, Joo H. Lee, Hongzhi Wu, Diego Gutierrez, and Min H. Kim. Simultane-ous acquisition of microscale reflectance and normals. In *ACM Transactions on Graphics (TOG), (Proc. SIGGRAPH Asia 2016)*, Vol. 35, No. 6, 2016.

[12] Giljoo Nam, Joo Ho Lee, Diego Gutierrez, and Min H. Kim. 2018. Practical SVBRDF Acquisition of 3D Objects with Unstructured Flash Photography. In *ACM Trans. Graph.*, 37, 6, Article 267 (November 2018), 12 pages. `https://doi.org/10.1145/3272127.3275017`.

[13] Jaesik Park, Sudipta Sinha, Yasuyuki Matsushita, Yu W. Tai, and In S. Kweon. Robust multiview photometric stereo using planar mesh parameterization. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016.

[14] Jérémy Riviere, Pieter Peers, and Abhijeet Ghosh. Mobile surface reflectometry. In *Computer Graphics Forum (2015)*, 2015.

[15] Borom Tunwattanapong, Graham Fyffe, Paul Graham, Jay Busch, Xurming Yu, Abhijeet Ghosh, and Paul Debevec. Acquiring reflectance and shape from continuous spherical harmonic illumination. In *ACM Transactions on Graphics (TOG)*, Vol. 32, No. 4, Article 109, 2013.

[16] Rui Xia, Yue Dong, Pieter Peers, and Xin. Tong. Recovering shape and spatially-varying surface reflectance under unknown illumination. In *Transactions on Graphics (TOG)*, Vol. 35, No. 6, Article 187, 2016.

[17] Zhiming Zhou, Guojun Chen, Yue Dong, David Wipf, Yong Yu, John Snyder, and Xin Tong. Sparse-as-possible svbrdf acquisition. In *ACM Transactions on Graphics (TOG)*, vol. 35, No. 6, Article 189, 2016.

[18] Michael Zollhöfer, Angela Dai, Matthias Innmann, Chenglei Wu, Marc Stamminger, Christian Theobalt, and Matthias Nießner. Shading-based refinement on volumetric signed distance functions. In *ACM Transactions on Graphics (TOG)*, Vol. 34, No. 4, Article 96, 2015.