# Fabric Appearance Benchmark

S. Merzbach and R. Klein

University of Bonn, Institute of Computer Science II, Germany

**Abstract**

*Appearance modeling is a difficult problem that still receives considerable attention from the graphics and vision communities. Though recent years have brought a growing number of high-quality material databases that have sparked new research, there is a general lack of evaluation benchmarks for performance assessment and fair comparisons between competing works. We therefore release a new dataset and pose a public challenge that will enable standardized evaluations. For this we measured 56 fabric samples with a commercial appearance scanner. We publish the resulting calibrated HDR images, along with baseline SVBRDF fits. The challenge is to recreate, under known light and view sampling, the appearance of a subset of unseen images. User submissions will be automatically evaluated and ranked by a set of standard image metrics.*

**CCS Concepts**

• ***Computing methodologies*** → ***Reflectance modeling; Appearance and texture representations;***

## 1. Introduction

During the last years, research on appearance modeling has gained significant attention. Thanks to new devices, which allow for mass acquisition, there is an abundance of material datasets. New software tools helped provide a plethora of artist-created material assets, two prominent examples being Substance Designer and Quixel Mixer. These advances have sparked a new line of deep learning approaches, which are trained on large-scale material databases, see Dong's survey [Don19]. Such databases exist as commercial products like Megascans, Poliigon, or Substance3D, but also as donation-based projects like Texturehaven, or as data releases of research projects [DAD*18; DJ18; LXR*18; MHRK19].

These developments have opened up many new research opportunities. However, there are virtually no standard benchmarks to assess the quality of reflectance models. This lack makes it difficult to fairly evaluate new works. For meaningful comparisons, it is either necessary to run new methods on several previously used datasets, many of which are not publicly available, or to re-implement existing works. The latter option is often ruled out due to high re-implementation effort or lack of data. To overcome this problem, we propose a new benchmark for appearance modeling that we pose as a public challenge to further motivate participation.

## 2. Related Work

Over the last years, several works released datasets of homogeneous BRDFs, either obtained from calibrated setups [MPBM03; FV14; DJ18], real-world photographs [BUSB13], or synthetically rendered under complex illumination [KGT*17; MMZ*18]. Though these datasets pose challenging problems, they are tailored to in-the-wild settings, where illumination estimation plays an equally if not even more important role than the material modeling itself. More important, most real surfaces are heterogeneous, thus limiting the use of such datasets for training or testing more complex models. Though more realistic datasets [DAD*18] feature hundreds of thousands of images of complex appearance, they are still based on renderings, i.e. there are no real-world reference images for comparison purposes. Such images do exist in datasets published with works on image-based material measurements, e.g. as Bidirectional Texture Functions (BTFs) [WGK14; FKH*18]. However, to the best of our knowledge, these materials were always fully released, i.e. there are no holdout images that would allow a fair evaluation in a competition setting as we propose here.

## 3. Challenge Statement

We now propose our challenge in detail. The main goal is to develop models that can be evaluated under given light and view directions to recreate observed pixel colors. To test model generalization, we provide only the directional sampling instead of some of the measurement images, and challenge participants to upload their model outputs for comparison against the unseen measurements.

The main goal is the highest possible reconstruction quality, i.e. model outputs should match the measurements as closely as possible. Models should furthermore be compact and efficient to evaluate, as they are to be used for rendering. Therefore, solving the problem with big neural networks is discouraged, but not explicitly forbidden, to open up potential new research directions. These two additional requirements will not be automatically evaluated, as we only allow data and no code submissions. As an example, most SVBRDF models fulfill the above requirements.

In a second branch of the challenge, the most dominant effects that are caused by mesoscopic surface structures, specifically shadowing and masking, are explicitly excluded from the evaluation. Please see Section 5.1 for details. We propose this separation to emphasize a model's ability to capture the *local* reflectance, without overimposed errors due to non-local shadowing or masking effects.

Details and instructions can be found on the challenge website: https://cg.cs.uni-bonn.de/appbench.

## 4. Dataset

We extend the existing UBOFAB19 dataset [MHRK19], which is based on measurements with the X-Rite TAC7. We publicly release two different extensions together with the proposed challenge:

**New SVBRDFs:** First, we provide new fits for all UBOFAB19 SVBRDFs, which so far only had *spatially uniform* Fresnel coefficients. The new SVBRDFs will have spatially varying Fresnel coefficients, and are fit with the latest Pantora version with better quality. This improves usability of the training data, as the uniform Fresnel coefficients could lead to ambiguous training samples for methods trained to directly regress to the SVBRDF parameters.

**New materials:** Second, we release 56 completely new materials for evaluating challenge submissions. Like in UBOFAB19, those materials are provided as radiometrically and geometrically calibrated HDR images and Pantora SVBRDF fits. See here for details.

**Holdout set:** Different from the UBOFAB19 release, which provides *all* TAC7 measurements for each material, we select 10% of the point-lit HDR images for the evaluation. Participants can only access the directional sampling of these images to reproduce their pixels' colors with their models. The automatic quality assessment for the uploaded renderings is described in the next section.

## 5. Benchmark Evaluation

We use the following metrics to judge the reconstruction quality:

- mean absolute deviation (MAD) & mean square error (MSE)
- Structural Similarity Index (SSIM) [WBSS04]
- CIE $\Delta E_{2000}$ [CIE01]
- HDR-VDP 2.2 [NMSC15]

We evaluate each metric on pairs of unseen measurement and the corresponding rerendered submission image. The error values are averaged over all pixels, images and materials in the holdout set.

### 5.1. Weighted evaluation

The measured materials are not perfectly planar. Therefore, some pixels are occluded or shadowed by other surface structures for some of the cameras and light sources. With known surface geometry, such pixels can be found by raytracing. In this challenge branch, we want to exclude such effects by weight maps that assign values between 0 and 1 to each pixel, and mask out occluded or shadowed areas. The weight maps $w$ are applied before error map computation. The error under metric $M$ is the sum over the error map pixels $i$, followed by a normalization by the weights:

$$E_M = \frac{\sum_i M(w \odot I, w \odot \hat{I})_i}{\sum_i w_i}, \; w_i = \min\left(m_i, \max(0, \langle \mathbf{n}_i, \mathbf{l}_i \rangle \cdot \langle \mathbf{n}_i, \mathbf{v}_i \rangle)\right),$$

where $\odot$ denotes the pixel-wise product between weight map $w$ and the images, and the masking $m_i = 0$ indicates that pixel $i$ is shadowed or occluded, otherwise $m_i = 1$. The weights in unmasked regions are the multiplied dot products respectively between surface normal and light or view direction. This weighting scheme has been applied before to fitting of reflectance data [BSN16] or as part of a perceptual BRDF similarity metric [Rym18]. It has the desirable property of weighing down pixels that have higher uncertainties.

## 6. Baseline Comparison

We use Pantora SVBRDF fits as a baseline representation, which we consider a reasonable choice, as they already reproduce the reflectance at a high quality. At the same time, they are certainly one of the simplest possible models to achieve this quality, as each pixel is represented only by a single lobe. More complex models are expected to increase the quality.

## 7. Running Time

We open the challenge to the public with publication of this poster. It will be open-ended and can easily be extended in the future with more materials or even by different material classes.

## 8. Acknowledgements

We cordially thank the X-Rite Bonn team for helpful discussions and their continuous support.

## References

[BSN16] BAGHER, M.M., SNYDER, J., and NOWROUZEZAHRAI, D. "A non-parametric factor microfacet model for isotropic brdfs". *ACM TOG* 35.5 (2016), 1–16 2.

[BUSB13] BELL, S., UPCHURCH, P., SNAVELY, N., and BALA, K. "OpenSurfaces: A richly annotated catalog of surface appearance". *ACM TOG* 32.4 (2013), 1–17 1.

[CIE01] CIE, ICOI. *Improvement to industrial colour-difference evaluation*. 2001 2.

[DAD*18] DESCHAINTRE, V., AITTALA, M., DURAND, F., et al. "Single-image svbrdf capture with a rendering-aware deep network". *ACM Transactions on Graphics (ToG)* 37.4 (2018), 1–15 1.

[DJ18] DUPUY, J. and JAKOB, W. "An adaptive parameterization for efficient material acquisition and rendering". *ACM TOG* 37.6 (2018) 1.

[Don19] DONG, YUE. "Deep appearance modeling: A survey". *Visual Informatics* (2019) 1.

[FKH*18] FILIP, J., KOLAFOVÁ, M., HAVLÍČEK, M., et al. "Evaluating Physical and Rendered Material Appearance". *The Visual Computer (Computer Graphics International 2018)* (2018) 1.

[FV14] FILIP, J. and VÁVRA, R. "Template-Based Sampling of Anisotropic BRDFs". *CGF* 33.7 (2014) 1.

[KGT*17] KIM, K., GU, J., TYREE, S., et al. "A lightweight approach for on-the-fly reflectance estimation". *Proceedings of ICCV*. 2017 1.

[LXR*18] LI, Z., XU, Z., RAMAMOORTHI, R., et al. "Learning to reconstruct shape and spatially-varying reflectance from a single image". *SIGGRAPH Asia 2018 Technical Papers*. ACM. 2018, 269 1.

[MHRK19] MERZBACH, S., HERMANN, M., RUMP, M., and KLEIN, R. "Learned fitting of spatially varying brdfs". *CGF*. Vol. 38. 4. 2019 1, 2.

[MMZ*18] MEKA, A., MAXIMOV, M., ZOLLHOEFER, M., et al. "LIME: Live Intrinsic Material Estimation". *Proceedings of CVPR*. 2018 1.

[MPBM03] MATUSIK, W., PFISTER, H., BRAND, M., and MCMILLAN, L. "A Data-Driven Reflectance Model". *ACM Transactions on Graphics* 22.3 (July 2003), 759–769 1.

[NMSC15] NARWARIA, M., MANTIUK, R., SILVA, M.P. DA, and CALLET, P. LE. "HDR-VDP-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images". *Journal of Electronic Imaging* 24.1 (2015), 010501 2.

[Rym18] RYMAN, D. *A Metric for Perceptual Distancebetween Bidirectional ReflectanceDistribution Functions*. 2018 2.

[WBSS04] WANG, Z., BOVIK, A.C., SHEIKH, H.R., and SIMONCELLI, E.P. "Image quality assessment: from error visibility to structural similarity". *IEEE transactions on image processing* 13.4 (2004), 600–612 2.

[WGK14] WEINMANN, M., GALL, J., and KLEIN, R. "Material Classification Based on Training Data Synthesized Using a BTF Database". *Proceedings of ECCV*. Springer International Publishing, 2014, 156–171 1.