Workshop on Material Appearance Modeling (2020)
H. Rushmeier and R. Klein (Editors)

Bonn Appearance Benchmark

S. Merzbach and R. Klein

University of Bonn, Germany

Abstract

There is a general shortage of standardized comparisons in the field of appearance modeling. We therefore introduce a benchmark for assessing the quality of reflectance models on a dataset of high quality material measurements obtained with a commercial appearance scanner. The dataset currently consists of 56 fabric materials which are measured as radiometrically calibrated HDR images together with a precise surface geometry. We pose a public challenge to attract further participation and spark new research. Participants evaluate their models on provided directional light and view sampling to recreate the appearance of a set of unseen images per material. The results are automatically evaluated under various image metrics and ranked in a public leaderboard. Our benchmark provides standardized testing and thus enables fair comparisons between related works. We also release baseline SVBRDF material fits.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture

1. Introduction

Appearance modeling deals with finding representations of captured or simulated reflectance. The goal is to recreate realistic reflectance behavior when evaluating the developed models. Before modelling, one needs to acquire real surface reflectance. This usually requires expensive and carefully calibrated setups that densely sample the surface in the angular domain of light and view directions. Specialized, self-calibrating hardware allows to capture spatially resolved reflectance on a large scale [Deb12, KNRS13, SRT*14, XR18]. In parallel developed software packages like Adobe's Substance Designer [Sub20a] and Quixel's Mixer [Mix20] provide convenient tools for artists to create realistic materials. Thus, obtaining databases of hundreds of different material samples is easily possible. Commercial databases like Adobe Substance3D [Sub20b], Poliigon Textures [Pol20], or the Quixel Megascans [Meg20] are abundant, but difficult to use for research purposes due to high purchase costs. A line of works [NDM05, DAD*18, DJ18, MHRK19] publicly released their datasets, enabling usage of such material collections as training data for machine learning techniques [LXR*18, LSC18, YLD*18, BMS*19, BL19, DAD*19, GLD*19, KXH*19, RJGW19, VCGLM19, BXS*20, BJK*20, RGJW20].

The growth of available data has so far not provided any means for systematic comparisons between existing works. Appearance modelling has the advantage that methods can be directly evaluated on training materials by just picking a new combination of light and view directions that is unseen in the training data. This is in fact how most works are validated. However, only few of them provide extensive comparisons with related works. This is mainly for two reasons: First, not all existing works provide their code, while the re-implementation effort is usually not justified for comparison purposes only. Second, a lot of works do not provide the datasets they use for their evaluation, ruling out comparisons of new methods against their previous results.

We are therefore motivated to improve this situation by establishing a benchmark for appearance modelling. We build upon the UBOFAB19 material database [MHRK19] which consists of 378 fabric materials, scanned with a commercial appearance scanner [XR18]. Each scan consist of several hundred radiometrically calibrated HDR images, associated with pixel-wise light and view directions. We extend this dataset with the *APPBENCH* release of 56 completely new fabrics. Contrary to UBOFAB19, 10% of the HDR images are held back for usage in the benchmark evaluation. We only provide the directional sampling for the images in this holdout set. Thus, one can ensure that comparisons are

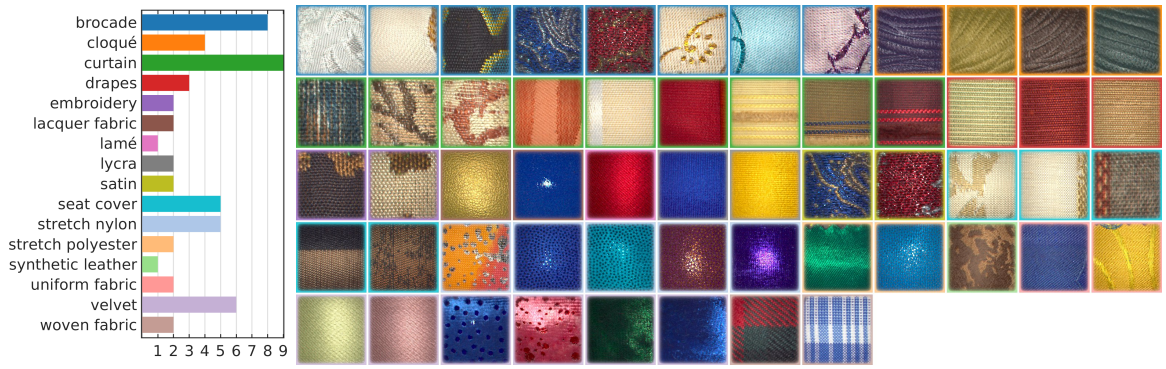

Figure 1: *Overview of the 56 new fabric materials released in our APPBENCH dataset, available at* https://cg.cs.uni-bonn.de/appbench/. *Colored frames indicate the fabric type.*

fair, as the images in the holdout set are guaranteed to be unseen. We pose the benchmark as a challenge on the codalab platform, where participants can upload their methods' reconstructions. These uploaded images are automatically evaluated against the ones in the holdout set under a set of standard image metrics.

2. Related Work

Perhaps most closely related to our work is the SynBRDF benchmark dataset [KGT*17]. It consists of 5000 Ward BRDFs, rendered on 5000 shapes under 20 natural environment maps. The authors create half a million LDR and HDR RGBD images together with ground truth information of BRDF parameters, 3D shape, illumination and camera pose. However, the dataset is more interesting for "in-the-wild" settings, whereas our contribution is based on radiometrically calibrated reflectance measurements. This allows much better investigation of the *reflectance* models' accuracy, as other error sources, resulting e.g. from inaccuracies in the illumination, are ruled out. More importantly, the BRDF parameters in SynBRDF are sampled from OpenSurfaces [BUSB13], a dataset of real-world photographs annotated with homogeneous BRDF parameters. However, variations in surface reflectance make up a crucial part of real object appearance and perfectly homoegenous surfaces are the exception. There are several other works with homogeneous BRDF datasets, most of which are measured in calibrated setups like gonioreflectometers [MPBM03, FVH14, FV14, DJ18]. We consider heterogenous surfaces a much more challenging and realistic setting, which is why we use spatially varying reflectance in our benchmark.

Deschaintre et al. [DAD*18] released an extensive dataset of renderings of spatially varying, but synthetic materials. Though considerably realistic, these materials tend to lack the last bit of imperfections and variations that distinguish them from real-world images. Furthermore, the realism of renderings is limited by the expressiveness of the BRDF model used during rendering, so other datasets like the synthetic renderings of measured Adobe Stock SVBRDFs [XSHR18] face the same limitation.

There are other image-based reflectance models like Bidirectional Texture Functions (BTFs) [DGNK97] which do not suffer from these limitations. Instead of BRDFs, they are composed of *apparent* BRDFs (ABRDFs), which contain arbitrarily complex reflectance, including interereflections and shadowing effects. Though they are available in several rich material databases [SSK03, RSK10, HM12, WGK14, FKH*18], BTFs are not suitable for arbitrarily glossy materials, as their acquisition, storage, post-processing and rendering grow more expensive with high glossiness due to the necessary denser angular sampling to faithfully capture all highlights.

The images in our dataset contain reflectance measurements of real-world materials. They are obtained with the commercial TAC7 appearance scanner, manufactured by X-Rite [XR18]. Similar to BTFs, our images contain ABRDFs, but with per pixel varying angular sampling. This avoids the error-prone step of resampling the measurement images. Our benchmark extends the existing UBOFAB19 dataset [MHRK19] of TAC7 fabric scans.

3. Dataset

In this section we describe the details of our material dataset. It consists of 56 new fabric materials that are selected from a wide range of fabric categories, ranging from brocades to velvet, see Fig. 1.

The TAC7 appearance scanner rotates material samples on a turntable under four fixed panchromatic cameras, 29 fixed white point-like LED light sources, as well as a strip-like light source (called linear light source, LLS) that is mounted on an arm and can be rotated to arbitrary inclination angles. Optionally, samples can be placed on a back-lit diffuser plate for translucency measurements. The surface geometry is ob-

tained via structured light measurements from a single projector. Color information is captured by rotating filter wheels in front of five of the LEDs. Measurements are performed by rotating the turntable to one of five orientations (0, 45, 90, 135 and 180 degrees). All four cameras then capture exposure series for each individual light source (and optionally color filter), including the structured light projector. Depending on a user-selected glossiness preset, the LLS is rotated with a step size of 6 (low gloss), 3 or 0.5 degrees (high gloss) for each turntable orientation. The total number of measured images (excluding structured light and back-lit images) per material are 388 point-lit, out of which 100 are color images, as well as respectively 280, 560 or 3300 line-lit images for the low-, medium- or high-gloss presets.

All images are radiometrically calibrated, i.e. all camera non-linearities are calibrated out of the data during HDR combination. The same holds for illumination or camera effects like light falloff or lens vignetting, which are removed via white-frame calibration. Color images are provided in linear sRGB color space under equal energy illuminant E.

Confidence maps: During post-processing, per-pixel confidence maps are calculated for each measurement image, indicating geometric and radiometric uncertainties in the data:

$$w_i = \min\left(m_i, \max(0, \langle \mathbf{n}_i, \mathbf{l}_i \rangle \cdot \langle \mathbf{n}_i, \mathbf{v}_i \rangle)\right),$$

where the masking term $m_i = 0$ indicates that pixel i is shadowed or occluded, otherwise $m_i \in [0,1]$ indicates potential radiometric uncertainties (e.g. over-exposure remaining after HDR combination for extremely bright highlights). The weights in unmasked regions are the product of the cosines of the angles between light respectively view and the normal. Thus, lower weights are assigned for grazing angles due to increased geometrical uncertainties. This weighting scheme is known from BRDF fitting [BSN16] or as part of a perceptual BRDF similarity metric [Rym18].

Holdout set: For each material, we select about 10% (40 images) of the point-lit images for a non-public test set, which is reserved for evaluation of challenge submissions. The images are selected to cover a wide range of light and view directions and are the same for each material.

4. Challenge

We pose the benchmark as a competition on the codalab platform. Participants can upload their result images, which are created by evaluating reflectance models, e.g. SVBRDFs, on the directional sampling corresponding to the images in the holdout set. The challenge is split into two branches:

Standard branch: The results are automatically compared against the holdout images using the following image metrics: mean absolute deviation (MAD), mean square error (MSE), Structural Similarity Index (SSIM) [WBSS04], HDR-VDP 2.2 [NMSC15], and for color images additionally CIE ΔE_{2000} [CIE01]. These metrics all return error maps, which are averaged over all pixels and, where available, over the color channels. Before computing the metrics, we apply masks that effectively correspond to a binarized version of the masking term m_i from above, i.e. indicating occluded or shadowed regionns in the images. The final metric scores are obtained by averaging over the 40 images times 56 materials.

Weighted branch: Here we additionally apply the confidence maps before computing the metrics, i.e. for a given metric we compute the error score E as

$$E = \frac{\sum_i M(w \odot I, w \odot \hat{I})_i}{\sum_i w_i},$$

where $\odot$ is the element-wise product, I is a reference image from the holdout set and $\hat{I}$ is the corresponding user reconstruction. $M(\cdot)$ is the channel-averaged error map under the selected metric, which is summed up over all pixels i and normalized by the sum over all confidence map pixels w_i.

We motivate this weighted branch as follows: The observations in the TAC7 images are ABRDFs, i.e. reflectance overlaid by shadowing, masking and interreflections. Furthermore, imperfections in the geometry reconstruction or fuzzy structures like small fibers can cause interactions between neighboring pixels. The confidence maps encode such uncertainties, at least to a limited degree. By weighing down pixels with lower confidence, we intend to put more focus on the actual reflectance behavior. In this way we hope to better assess the quality of purely local reflectance models like BRDFs that cannot represent shadows or other global effects.

Baselines: We use single-lobe Ward SVBRDFs fit with the Pantora software [XR20] as baseline model.

Duration: There is no time-limit for the challenge. We plan to extend it by further branches covering new material classes in the future.

5. Acknowledgements

We thank the X-Rite Bonn team for their continous support.

References

[BJK*20] Boss M., Jampani V., Kim K., Lensch H., Kautz J.: Two-shot spatially-varying brdf and shape estimation. In *Proceedings of CVPR* (2020), pp. 3982–3991. 1

[BL19] Boss M., Lensch H.: Single image brdf parameter estimation with a conditional adversarial network. *arXiv preprint arXiv:1910.05148* (2019). 1

[BMS*19] Bode L., Merzbach S., Stotko P., Weinmann M., Klein R.: Real-time multi-material reflectance reconstruction for large-scale scenes under uncontrolled illumination from rgb-d image sequences. In *3DV* (2019), IEEE, pp. 709–718. 1

[BSN16] Bagher M., Snyder J., Nowrouzezahrai D.: A non-parametric factor microfacet model for isotropic brdfs. *ACM TOG 35*, 5 (2016), 1–16. 3

[BUSB13] BELL S., UPCHURCH P., SNAVELY N., BALA K.: Opensurfaces: A richly annotated catalog of surface appearance. *ACM TOG 32*, 4 (2013), 1–17. 2

[BXS*20] BI S., XU Z., SUNKAVALLI K., KRIEGMAN D., RAMAMOORTHI R.: Deep 3d capture: Geometry and reflectance from sparse multi-view images. In *Proceedings of CVPR* (2020), pp. 5960–5969. 1

[CIE01] CIE I.: Improvement to industrial colour-difference evaluation, 2001. 3

[DAD*18] DESCHAINTRE V., AITTALA M., DURAND F., DRETTAKIS G., BOUSSEAU A.: Single-image svbrdf capture with a rendering-aware deep network. *ACM TOG 37*, 4 (2018). 1, 2

[DAD*19] DESCHAINTRE V., AITTALA M., DURAND F., DRETTAKIS G., BOUSSEAU A.: Flexible svbrdf capture with a multi-image deep network. *CGF 38*, 4 (July 2019). 1

[Deb12] DEBEVEC P.: The light stages and their applications to photoreal digital actors. *SIGGRAPH Asia 2*, 4 (2012). 1

[DGNK97] DANA K., GINNEKEN B., NAYAR S., KOENDERINK J.: Reflectance and Texture of Real World Surfaces. In *CVPR* (Jun 1997), pp. 151–157. 2

[DJ18] DUPUY J., JAKOB W.: An adaptive parameterization for efficient material acquisition and rendering. *ACM TOG 37*, 6 (2018). 1, 2

[FKH*18] FILIP J., KOLAFOVÁ M., HAVLÍČEK M., VÁVRA R., HAINDL M., H. R.: Evaluating Physical and Rendered Material Appearance. *TVC* (2018). 2

[FV14] FILIP J., VÁVRA R.: Template-based sampling of anisotropic BRDFs. *CGF 33*, 7 (2014). 2

[FVH14] FILIP J., VAVRA R., HAVLICEK M.: Effective acquisition of dense anisotropic brdf. In *ICPR* (2014). 2

[GLD*19] GAO D., LI X., DONG Y., PEERS P., XU K., TONG X.: Deep inverse rendering for high-resolution svbrdf estimation from an arbitrary number of images. *ACM TOG 38*, 4 (2019). 1

[HM12] HAINDL M. FILIP J. V. R.: Digital material appearance: the curse of tera-bytes. *ERCIM News*, 90 (2012), 49–50. 2

[KGT*17] KIM K., GU J., TYREE S., MOLCHANOV P., NIESSNER M., KAUTZ J.: A lightweight approach for on-the-fly reflectance estimation. In *Proceedings of ICCV* (2017). 2

[KNRS13] KÖHLER J., NÖLL T., REIS G., STRICKER D.: A full-spherical device for simultaneous geometry and reflectance acquisition. In *WACV* (2013), IEEE, pp. 355–362. 1

[KXH*19] KANG K., XIE C., HE C., YI M., GU M., CHEN Z., ZHOU K., WU H.: Learning efficient illumination multiplexing for joint capture of reflectance and shape. *ACM TOG 38*, 6 (2019), 1–12. 1

[LSC18] LI Z., SUNKAVALLI K., CHANDRAKER M.: Materials for masses: Svbrdf acquisition with a single mobile phone image. In *Proceedings of ECCV* (2018), pp. 72–87. 1

[LXR*18] LI Z., XU Z., RAMAMOORTHI R., SUNKAVALLI K., CHANDRAKER M.: Learning to reconstruct shape and spatially-varying reflectance from a single image. In *SIGGRAPH Asia* (2018), ACM, p. 269. 1

[Meg20] MEGASCANS: Quixel, 2020. URL: https://quixel.com/megascans/. 1

[MHRK19] MERZBACH S., HERMANN M., RUMP M., KLEIN R.: Learned fitting of spatially varying brdfs. In *CGF* (2019), vol. 38. 1, 2

[Mix20] MIXER: Quixel, 2020. URL: https://quixel.com/mixer. 1

[MPBM03] MATUSIK W., PFISTER H., BRAND M., MCMILLAN L.: A data-driven reflectance model. *ACM TOG 22*, 3 (July 2003), 759–769. 2

[NDM05] NGAN A., DURAND F., MATUSIK W.: Experimental analysis of brdf models. *Rendering Techniques 2005*, 16th (2005), 2. 1

[NMSC15] NARWARIA M., MANTIUK R., SILVA M. D., CALLET P. L.: 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. 3

[Pol20] POLIIGON: Poliigon Textures, 2020. URL: https://www.poliigon.com/search?type=texture. 1

[RGJW20] RAINER G., GHOSH A., JAKOB W., WEYRICH T.: Unified neural encoding of btfs. In *CGF* (2020), vol. 39, Eurographics Association, pp. 1–13. 1

[RJGW19] RAINER G., JAKOB W., GHOSH A., WEYRICH T.: Neural btf compression and interpolation. *CGF 38*, 2 (Mar. 2019). 1

[RSK10] RUMP M., SARLETTE R., KLEIN R.: Groundtruth data for multispectral bidirectional texture functions. In *CGIV 2010* (June 2010), Society for Imaging Science and Technology, pp. 326–330. 2

[Rym18] RYMAN D.: A metric for perceptual distance between bidirectional reflectance distribution functions, 2018. 3

[SRT*14] SANTOS P., RITZ M., TAUSCH R., SCHMEDT H., MONROY R., STEFANO A. D., POSNIAK O., FUHRMANN C., FELLNER D.: Cultlab3d: On the verge of 3d mass digitization. In *Eurographics Workshop on GCH* (2014), pp. 65–73. 1

[SSK03] SATTLER M., SARLETTE R., KLEIN R.: Efficient and realistic visualization of cloth. In *EGSR* (June 2003). 2

[Sub20a] SUBSTANCE DESIGNER: Adobe, 2020. URL: https://www.substance3d.com/products/substance-designer/. 1

[Sub20b] SUBSTANCE3D SHARE: Adobe, 2020. URL: https://share.substance3d.com/. 1

[VCGLM19] VIDAURRE R., CASAS D., GARCES E., LOPEZ-MORENO J.: Brdf estimation of complex materials with nested learning. In *IEEE WACV* (2019), IEEE, pp. 1347–1356. 1

[WBSS04] WANG Z., BOVIK A., SHEIKH H., SIMONCELLI E.: Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing 13*, 4 (2004), 600–612. 3

[WGK14] WEINMANN M., GALL J., KLEIN R.: Material classification based on training data synthesized using a btf database. In *Proceedings of ECCV* (2014), Springer International Publishing, pp. 156–171. 2

[XR18] X-RITE: Tac7-scanner, June 2018. URL: http://web.archive.org/web/20180615015942/https://www.xrite.com/categories/appearance/tac7. 1, 2

[XR20] X-RITE: Pantora material hub, Apr. 2020. URL: https://web.archive.org/web/20190831131500/https://www.xrite.com/categories/appearance/pantora-software. 3

[XSHR18] XU Z., SUNKAVALLI K., HADAP S., RAMAMOORTHI R.: Deep image-based relighting from optimal sparse samples. *ACM TOG 37*, 4 (2018), 126. 2

[YLD*18] YE W., LI X., DONG Y., PEERS P., TONG X.: Single image surface appearance modeling with self-augmented cnns and inexact supervision. In *CGF* (2018), vol. 37, pp. 201–211. 1