Multiplexing and Multi-Mode Super-Resolution for Digital Material Appearance

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Dennis den Brok
aus
Kleve

Bonn, 11.10.2022

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn.

1. Gutachter:	Prof. Dr. Reinhard Klein
2. Gutachter:	Prof. Dr. Matthias Hullin
Tag der Promotion:	19.12.2022
Erscheinungsjahr:	2023

Abstract

The visual appearance of materials on the surface of objects contributes greatly to the overall perception of virtually every scene we encounter every day. However, complex material appearance can be neigh impossible to model mathematically. Consequently, it can be difficult to reproduce in computer-generated imagery. As a work-around, image-based digital representations such as the bidirectional texture function (BTF) treat material appearance as a "black box" which reproduce a given material's appearance under the desired conditions by appropriately interpolating between actual images of the material at hand. Effectively, the problem of physically-based modelling is replaced with the problem of measuring: in order to ensure high-quality, photo-realistic renderings, thousands to tens of thousands images of a given material under varying viewing and lighting conditions are required. The design of appearance acquisition devices inevitably involves choices with negative impact on acquisition speed and accuracy.

The present thesis introduces a number of methods which alleviate this impact and therefore make appearance acquisition much more practical. It is demonstrated that a generalization of phenomenological models usually used for the purpose of BTF compression allows for reconstruction of a full BTF measurement from a sparse measurement, greatly enhancing measurement speed and enabling cheaper and more compact acquisition devices.

As another way of reducing acquisition time, illumination patterns of multiple instead of single light sources are used, which significantly brightens materials with low albedo or lots of self-shadowing, allowing for shorter shutter times. It is shown that the shot noise present in the reconstruction of the desired images under single-light illumination can be effectively reduced using the same or similar phenomenological models.

Both methods seem orthogonal and thus applicable simultaneously. However, due to the dynamic range of BTF data, the employed models involve non-linear transformations which make a combination unfeasible. A novel model is introduced, along with an alternating least-squares algorithm to compute it efficiently, which deals with dynamic range in a different way and allows for simultaneously sparse and illumination-multiplexed acquisition.

As a means to improve the optical resolution of material measurements, or allow for larger material samples to be measured, a texture-space super-resolution algorithm based on deep-learning techniques is introduced. It is shown that despite working on an image-by-image basis, the resulting BTF does not exhibit obvious artifacts, and that data available realistically is sufficient ground truth for the purpose of training.

Contents

CHAPTER 1

Introduction

Computer-generated imagery (CGI), even in the narrow sense of imagery created from digital representations of 3D scenes, has changed dramatically over the course of only half a century. From the humble beginnings of simple, low-resolution, static shapes to a state-of-the-art which, at its best, can itself only be described as humbling (cf. Fig. 1.1). In fact, CGI can be so deceptively realistic nowadays that researchers in computer vision are developing methods to assist humans in telling measured data, i.e. photos or videos, and CGI apart [103]. Given these developments, it is not surprising that CGI has become an indispensable tool in many applications: in research, it can be used for visualization, for creating synthetic training data for machine learning algorithms [104], or for simulated scenes to be used in psychological studies [105]. Computer games would be entirely impossible without CGI, and many movies, TV shows and documentaries would look completely different if they had to rely on other animation techniques only. Finally, in commercial and industrial applications, CGI plays an increasingly important role, enabling more flexible advertising, product design, and so on.

The achievements of 3D computer graphics as a field of study cannot be attributed to a single sub-discipline – they have been reached through continuing efforts in fields as diverse as rendering, geometry processing and material appearance, each of which has seen similarly impressive developments in the course of the decades. While in all generality none of these disciplines can be singled out as the most important one, a case can be made for material appearance, because there exist image-based material appearance models which incorporate micro- and meso-scale surface geometry and thereby, as I shall explain in the subsequent section, lower the demands on both geometry processing and rendering.

Finally, note that the term "computer-generated imagery" is somewhat misleading. While CGI is indeed generated using computers, its creation often relies on data ultimately not generated by a computer, but obtained somehow from the real world. For instance, an artist may build an elaborate digital 3D scene consisting of virtual objects with complex surface properties, either from their imagination, infused by their experience of the real world, or directly inspired by or closely following an existing scene in the real world. But of course it is not always desirable or possible for a human to act as the computer's senses. Consequently, there are many ways of digitizing the physical properties like surface reflectance of objects

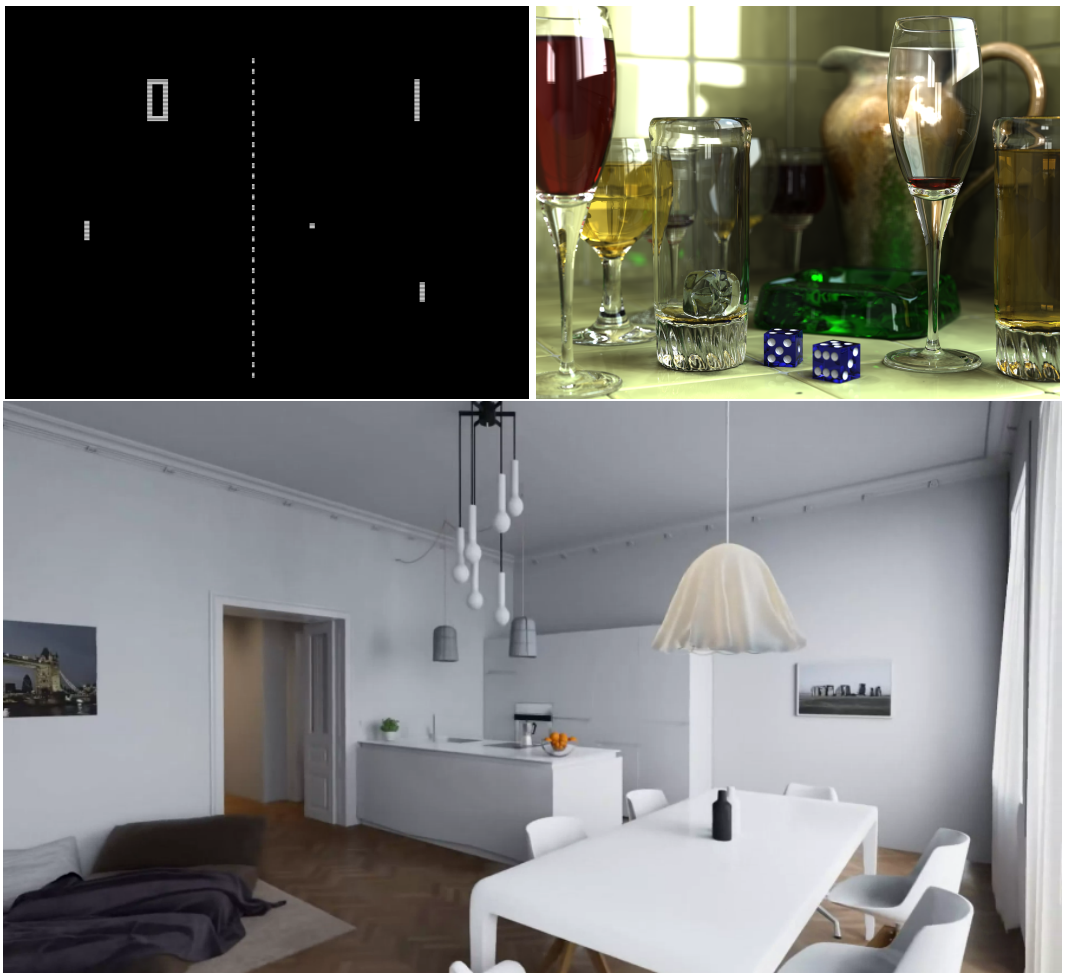


Figure 1.1: Top left: Screen photo of the famous arcade video game *Pong* (1972). Top right: Rendering of a scene created by an artist. Bottom: Screenshot of an interactive live architecture demo of a well-known 3D rendering engine, courtesy of `UE4Arch.com`. It is noteworthy that while photo-realistic, the above 3D scenes do not contain optically complex spatially-varying materials.

present in a scene. In the case of digitization of material appearance, virtually all sensing devices consist of digital cameras and rely on light sources within (LEDs, . . .) or outside of (environmental light) the operator's control. Additionally, image-based representations of material appearance require a certain amount of images of the material to be captured in order to avoid artifacts due to a low sampling rate along the angular dimensions, i.e. directions of incoming/outgoing light (cf. Fig. 1.2). Sensing devices for material appearance thus suffer both from the constraints imposed by the underlying hardware, e.g. optical resolution or brightness of light sources, and those imposed by how the hardware is organized, e.g. the number of cameras and light sources and their placement in space.

The purpose of the present thesis is to extend the range of available methods for capturing image-based material appearance, with a special focus on lifting the constraints imposed by

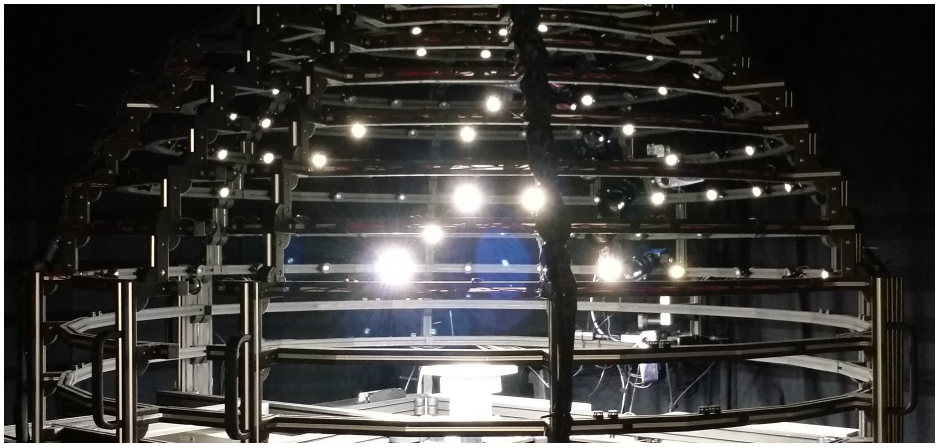


Figure 1.2: Acquisition device for image-based digital material appearance. Many light sources on a hemisphere above the material sample illuminate the scene, which is captured by a number of cameras.

the hardware. To this end, I introduce a number of methods which

- increase optical resolution, allowing for higher-quality digital material appearance or greater flexibility in the design of acquisition devices;
- increase angular resolution, enabling faster acquisition and more compact acquisition set-ups;
- increase the amount of light hitting the surface of the material to be measured, allowing for shorter shutter times and much faster measurements.

All of these methods make use of novel data-driven models derived from databases of measured material appearance.

In the subsequent sections, I provide definitions of the relevant terms, along with an overview of appearance models in the given sense, and an outline of the methods I have developed together with my co-authors.

1.1 Digital material appearance

The term “material appearance” is not very rigidly defined in general, and in the present thesis, it only serves as an umbrella term for a small number of concepts that are, possibly up to optional parameters. More generally, a material’s appearance may encompass properties other than visual appearance, such as its tactile or acoustic properties, or even changes over time [106]. However, in the following, “material appearance” is to be read as “visual appearance of a material, regardless of the object composed of the material”; that is, the way a particular material interacts with light.

The following subsections are loosely based on “Visual Texture” by Haindl and Filip [46] and course notes by Michael Weinmann [60].

1.1.1 Reflectance functions

Superficially, the interaction of light with a material can be described in mathematical terms by means of a *general reflectance function* [46]

$$f(\lambda_i, \mathbf{x}_i, t_i, \boldsymbol{\omega}_i, \lambda_o, \mathbf{x}_o, t_o, \boldsymbol{\omega}_o, \boldsymbol{\omega}_t) = \frac{\mathrm{d}L_o(\lambda_o, \mathbf{x}_o, t_o, \boldsymbol{\omega}_o, \boldsymbol{\omega}_t)}{L_i(\lambda_i, \mathbf{x}_i, t_i, \boldsymbol{\omega}_i)\,\langle \boldsymbol{\omega}_i, \mathbf{n}\rangle\,\mathrm{d}\boldsymbol{\omega}_i}.$$

Its value for a given set of parameters is the fraction of the spectral irradiance $L_i(\lambda_i, \mathbf{x}_i, t_i, \boldsymbol{\omega}_i)\,\langle \boldsymbol{\omega}_i, \mathbf{n}\rangle$ incident on the material's surface at point $\mathbf{x}_i \in \mathbf{R}^3$ in space with surface normal $\mathbf{n} \in \mathbf{R}^3$ and at time $t_i \in \mathbf{R}$ with wavelength $\lambda_i \in \mathbf{R}$ from direction $\boldsymbol{\omega}_i \in \mathbf{R}^3$ re-emitted as spectral radiance $L_o(\lambda_o, \mathbf{x}_o, t_o, \boldsymbol{\omega}_o, \boldsymbol{\omega}_t)$ from the material's surface at point $\mathbf{x}_o \in \mathbf{R}^3$ in space and at time $t_o \in \mathbf{R}$ with frequency $\lambda_o \in \mathbf{R}$ along direction $\boldsymbol{\omega}_o \in \mathbf{R}^3$ and direction $\boldsymbol{\omega}_t \in \mathbf{R}^3$ of transmission. In this general form, most optical effects which make up everyday materials' appearance can be adequately represented, such as specular and diffuse reflection, fluorescence, phosphorescence, sub-surface scattering, transmission, and even ageing. Given a general reflectance function, a digital image of an object the surface of which consists of the represented material's appearance can be generated using the *rendering equation* [107]:

$$L_o(\lambda, \mathbf{x}, \boldsymbol{\omega}_o, t) = L_e(\lambda, \mathbf{x}, \boldsymbol{\omega}_o, t) + \int_\Omega f(\lambda, \mathbf{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o, t)\, L_i(\lambda, \mathbf{x}, \boldsymbol{\omega}_i, t)\,\langle \boldsymbol{\omega}_i, \mathbf{n}\rangle\,\mathrm{d}\boldsymbol{\omega}_i,$$

where L_e denotes emission by the surface itself. Note that the reflectance function f in this version of the rendering equation, taken from the publication which introduced it to the field of computer graphics back in 1986 and, as of this writing, still one of the most commonly used forms, has less parameters than the general reflectance function introduced above. This is because the latter's generality comes with severe drawbacks which render general reflectance functions extremely impractical to deal with and thus seldomly-used: even under simplifying assumptions on the underlying physics, evaluating a general reflectance function means solving the problem of global illumination for a specific set of parameters, possibly including scattering between points on the surface, which is recursive. It is therefore extremely difficult, if not impossible, to write down closed-form general reflectance functions for any but classes of comparatively simple materials. Consequently, evaluating and particularly fitting general reflectance functions to real-world material samples can become arbitrarily expensive. For this reason, these *analytical* models are often replaced by *phenomenological* models, where physical soundness is given up in favor of much simpler models based on interpolation of measured data. However, as can be inferred from the number of parameters to general reflectance functions, the necessary data can be enormously cumbersome to deal with: depending on the quality of the interpolation method and the complexity of the material's reflectance, the high-dimensional parameter space needs to be sampled very densely, resulting in prohibitive amounts of data to be acquired, processed, stored, and interpolated when evaluating the function.

Fortunately, applications requiring a material's entire general reflection function are rare. Most of the time, it is in fact sufficient to restrict the function to a lower-dimensional parameter space as in the above rendering equation, e.g. by ignoring the time parameters $t_{\{i,o\}}$ and, thereby, effects like ageing and phosphorescence. Many variants of such restrictions have proven very useful in practice over the decades and have therefore been given names. In the following, I introduce the variants most relevant to the present thesis.

For all variants, it is assumed that the appearance of the materials to be represented

- is not phosphorescent; i.e. $f(t_i, t_o) = f(t) \cdot \delta(t_i, t_o)$;
- does not change over time; i.e. $f(t) \equiv \text{const}$;
- is not fluorescent; i.e. $f(\lambda_i, \lambda_o) = f(\lambda) \cdot \delta(\lambda_i, \lambda_o)$;
- does not exhibit transmission; i.e. we may omit the parameter $\boldsymbol{\omega}_t$.

We further assume that light takes either monochromatic (one-dimensional) or tri-stimulus values (three-dimensional, e.g. RGB) and, in conjunction with the previous assumptions, may therefore omit both the wavelength parameter λ and the time parameter t. For better readability, light is assumed to be monochromatic for the rest of this introduction. The generalization to the tri-stimulus case is straight-forward. Finally, in the above, $\mathbf{x}_{\{i,o\}} \in \mathbf{R}^3$ are arbitrary points in space; however, we shall assume that $\mathbf{x}_{\{i,o\}} \in \mathbf{R}^2$ are points on a two-dimensional surface supposed to approximate the material's proper surface.

Textures

Under the assumption that f is constant for fixed $\mathbf{x}_{\{i,o\}}$, and that either $f(\mathbf{x}_i, \mathbf{x}_o) = 0$ whenever $\mathbf{x}_i \neq \mathbf{x}_o$, we arrive at a function $f(\mathbf{x})$ of a single two-dimensional parameter corresponding to positions on the material's surface; that is, a *texture* as in classical 3D computer graphics texture mapping. Note that even though f does not depend on lighting or viewing directions, using the above rendering equation the material's appearance as seen by an observer varies with the direction of incoming light due to the cosine term $\langle \boldsymbol{\omega}_i, \mathbf{n} \rangle$. This so-called "Lambertian" reflectance occurs when a surface scatters light perfectly diffusely. Textures can thus be used to describe approximately diffuse materials, possibly with varying color and albedo across the surface, such as certain types of wallpapers and wood. Note that if the material's surface is not entirely flat, effects such as parallax and inter-reflections between points on the surface may occur, which may have to be treated separately at possibly great computational costs.

Bidirectional reflectance distribution functions

Conversely, if f is assumed constant for fixed $\boldsymbol{\omega}_{\{i,o\}}$, i.e. no variation occurs across the material's surface, we obtain a function $f(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o)$ of the directions of incoming and outgoing light, a *bidirectional reflectance distribution function* (BRDF) (cf. Fig. 1.3). They lend

themselves to describing materials with approximately no surface variation, such as various kinds of paints, plastics, metals, etc. As with textures, effects due to non-trivial surface geometry may have to be accounted for separately. Indeed, in conjunction with the above assumptions, a physically plausible BRDF is

- *non-negative*: $f(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o) \geq 0$;
- *Helmholtz reciprocal*: $f(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o) = f(\boldsymbol{\omega}_o, \boldsymbol{\omega}_i)$;
- *energy-conserving*: $\int_\Omega f(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o)\langle\boldsymbol{\omega}_o, \mathbf{n}\rangle \mathrm{d}\boldsymbol{\omega}_o \leq 1$ for all $\boldsymbol{\omega}_i \in \Omega$.

Spatially-varying BRDFs

Spatially-varying BRDFs (SVBRDFs) of the form $f(\mathbf{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o)$ are a natural generalization of both BRDFs and textures, in that as functions of $\mathbf{x}$ and $\boldsymbol{\omega}_{i,o}$, they take values in the spaces of BRDFs and textures, respectively. Consequently, they lend themselves to representations of material appearance that can locally be described by BRDFs, but not necessarily globally so, and where the surface geometry can be determined with a sufficient degree of accuracy. Examples include approximately smooth opaque materials such as wallpaper, polished wood, and fine-structured leathers and cloths. Whenever the material's appearance is locally not a BRDF, e.g. in the presence of subsurface scattering or inter-reflections, which may be observed on coarse-structured leathers and cloths, or the material's surface geometry is complex, for instance in the case of many types of carpets, more sophisticated material representations may allow for better results, albeit at the cost of SVBRDFs' simplicity. This is particularly true in real-time applications, where these effects are difficult to account for in an appropriate amount of time.

Bidirectional texture functions

Bidirectional texture functions (BTFs) of the form $f(\mathbf{x}, \boldsymbol{\omega}_i, \boldsymbol{\omega}_o)$, in turn, generalize SVBRDFs, even though they have the same signature as SVBRDFs, because as functions of $\vec{x}$, they do not take values in the space of BRDFs, but in that of *apparent* BRDFs (ABRDFs) (cf. Fig. 1.3). These have the same signature as BRDFs, but they are not assumed to be physically plausible BRDFs in the above sense. In particular, they are not constrained by conservation of energy. As a result, they inherently can account for effects which are non-local and may therefore violate conservation of energy locally:

This might occur, for instance, if a certain point on a material's surface receives light not only directly from the light source, but also indirectly from a different point on the material's surface (*inter-reflection*), or no light at all because the surrounding surface casts a shadow (*self-shadowing*). Similarly, light might be scattered underneath the material's surface (*subsurface scattering*). ABRDFs and, by extension, BTFs account for all of these effects. Note, however, that the BTF's signature demands that light sources are approximately infinitely distant and their light, hence, be directional: a BTF cannot reproduce cases where a

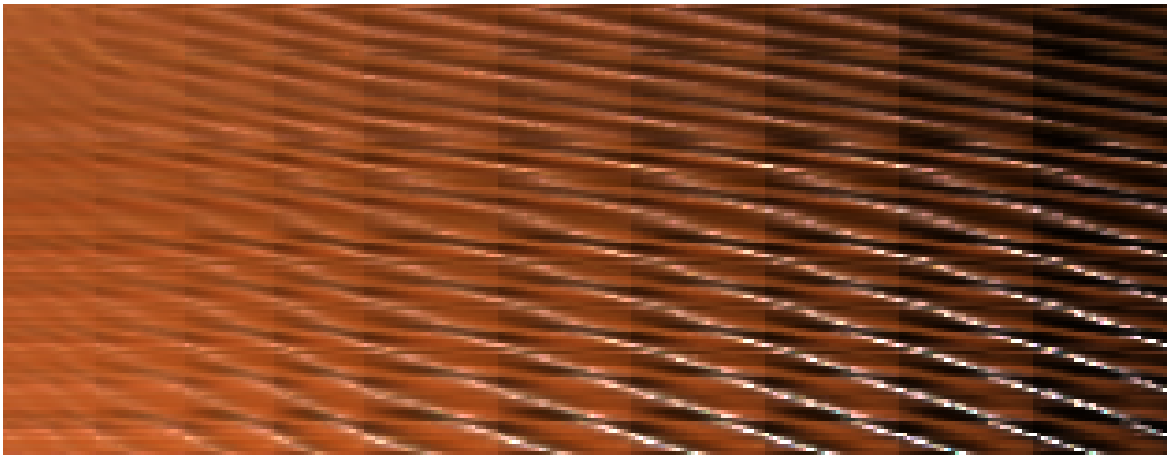


Figure 1.3: Example of a sampled ABRDF of a leather-like material. x-axis: sampled $\boldsymbol{\omega}_i$, y-axis: sampled $\boldsymbol{\omega}_o$.

ray of light hits the material's surface at a single point, or at multiple points but with varying angles, which would require a distinction between surface points of incoming and outgoing light $\mathbf{x}_i$ and $\mathbf{x}_o$ as in the case of the general reflection function described above.

Furthermore, in practice a material's surface geometry might only be known up to crude approximation, either by choice or due to the difficulty of capturing its complex structure. As a rather extreme example, one might not be able to determine a square patch of long pile carpet's exact surface geometry, so one might resort to representing it simply by a 2D square in 3D space. Due to parallax, the *actual* surface point $\mathbf{x}$ may thus vary greatly and therefore exhibit completely different optical properties for varying outgoing light directions ω_o. Again, for fixed $\mathbf{x}$, this means the function $f(\mathbf{x})$ may easily violate conservation of energy and therefore not be a proper BRDF, even if the material's appearance at the *actual* surface point could be represented by a BRDF.

In a sense, BTFs can thus be considered a "black box" which encapsulates all kinds of optical effects. BTFs and their modelling and acquisition are the present thesis' main focus.

1.1.2 Appearance models

A material's appearance is determined on several scales, from its subatomic structure over micro-scale features such as tiny flakes and fibres and meso-scale structures like weave patterns up to the macro-scale, where deviations of the surface geometry like curvature, dents, wrinkles and bumps may cast shadows or cause inter-reflections. While this is obviously infeasible to model precisely taking all scales into consideration, increasingly successful attempts have been made at modelling material appearance both analytically and phenomenologically on scales relevant for human perception.

Analytical models

Analytical models are typically based on idealizations of the underlying physical reality. This often leads to comparatively simple models with only small numbers of parameters, which makes them easy to deal with once the difficult step of estimating appropriate model parameters for a given collection of samples has been overcome. As an example, the anisotropic *Ward* model, introduced in 1992 and as of this writing still commonly used, makes certain assumptions about functions describing the height of possibly specular homogeneous material surfaces, giving rise to a Gaussian-shaped specular lobe of the form

$$f(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o) = \rho_s \frac{1}{\sqrt{\langle \boldsymbol{\omega}_i, \mathbf{n} \rangle \langle \boldsymbol{\omega}_o, \mathbf{n} \rangle}} \frac{\langle \boldsymbol{\omega}_i, \mathbf{n} \rangle}{4\pi\alpha_x\alpha_y} \exp\left[-2 \frac{\left(\frac{\langle \mathbf{x}, \mathbf{h} \rangle}{\alpha_x}\right)^2 + \left(\frac{\langle \mathbf{y}, \mathbf{h} \rangle}{\alpha_y}\right)^2}{1 + \langle \mathbf{h}, \mathbf{n} \rangle} \right],$$

with $\rho_s \in \mathbf{R}$ some weight, $\mathbf{x}, \mathbf{y} \in \mathbf{R}^3$ the directions of anisotropy with corresponding weights $\alpha_x, \alpha_y \in \mathbf{R}$, and $\mathbf{h} = \frac{\boldsymbol{\omega}_i + \boldsymbol{\omega}_o}{\|\boldsymbol{\omega}_i + \boldsymbol{\omega}_o\|_2}$ the half vector of $\boldsymbol{\omega}_i$ and $\boldsymbol{\omega}_o \in \mathbf{R}^3$ [108]. In order to achieve greater accuracy when modelling real-world materials, reflection functions are often constructed from several functions like the above, for instance as a sum of Ward functions with different parameters and a diffuse, e.g. Lambertian, reflectance function. Before and after the Ward model, a number of similar parametric BRDF models have been proposed, some of which are analytical models derived from the underlying physics, some of which are not [109, 110].

Phenomenological models

Phenomenological models do not make assumptions about the underlying physical reality as in the above and only aim at reproducing the phenomenon as faithfully as possible, often relying directly on measured data to reproduce material appearance, for instance by interpolating between measured data points using an appropriate function. This allows for accurate representation of material appearance of almost arbitrary complexity, constrained primarily by limitations of the acquisition setup and processing hardware. As an example, while general BRDFs are 4-dimensional functions (with $\boldsymbol{\omega}_{i,\,o}$ represented in spherical coordinates), *isotropic* BRDFs, the value of which is invariant under simultaneous rotation of $\boldsymbol{\omega}_{i,\,o}$ about the surface normal $\mathbf{n}$, only depend on three parameters after an appropriate change of variables (now known as the *Rusinkiewicz* parameterization) [57]. A measured isotropic BRDF can thus be represented as a 3-dimensional tensor and evaluated using barycentric coordinates for interpolation, given a sufficient sampling rate known to be achievable in practice [24]. Matusik et al. [23] further demonstrated that high-resolution measured isotropic BRDFs can be approximated well by a linear combination of a small number of basis BRDFs. Similarly, recent years have seen the development of phenomenological models derived from measured data using modern deep learning techniques [111].

Modelling spatially-varying appearance

Models for spatially-varying appearance like SVBRDFs and BTFs are at least partially phenomenological in general, because the spatial variance is typically extremely difficult to derive from first principles, let alone in a parametric way that allows for fitting model parameters to a real-world material sample. Indeed, SVBRDFs are commonly described simply as "images" comprised of model parameters, whereas BTFs are usually modelled based on entirely *image-based*, i.e. phenomenological, representations (with notable exceptions [112]):

For reasons which will become clear soon, a measured BTF is often re-sampled such that both the sets of BRDF- and texture-space sample coordinates are Cartesian products of finite subsets of Ω and $\mathbf{R}$, respectively. Let n_i, n_o, w and $h \in \mathbf{N}$ be the corresponding resolutions. Then the BTF can be thought of as a tensor $\mathcal{B} \in \mathbf{R}^{n_i \times n_o \times w \times h}$. After appropriate reshaping, it can be written as a matrix $\mathbf{B} \in \mathbf{R}^{n_i \cdot n_o \times w \cdot h}$; that is, its columns and rows represent the BTF's ABRDFs and bidirectional textures, respectively. In either form, the BTF can easily be evaluated for given coordinates $(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o, \mathbf{x})$

$$\mathbf{B}(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o, \mathbf{x}) = \sum_{i=1}^{3} \sum_{j=1}^{3} \lambda_{ij} \mathbf{B}(\boldsymbol{\omega}_i^{(i)}, \boldsymbol{\omega}_o^{(j)}, \mathbf{x}),$$

where λ_{ij} are interpolation weights and $\boldsymbol{\omega}_{i,o}^{(i,j)}$ the interpolated sampled directions of incoming and outgoing light, determined via 2D Delaunay triangulation of the corresponding sample points; $\mathbf{x}$ can be dealt with just as in regular texture mapping. Computing the weights is much more convenient when for all $\mathbf{x}$ the sampled $\boldsymbol{\omega}_{i\,o}$ are the same, as otherwise, there is a lot of computational or storage overhead involved. In particular, the interpolation weights can be pre-computed at a sufficient resolution, which makes it possible to render BTFs in real-time on a GPU, provided they fit in the GPU's memory.

However, a certain minimum BRDF-space sampling rate is required in order for the measured BTF to produce results close to the material's proper BTF, in particular for materials with a glossy and/or uneven surface, which may lead to unpleasant interpolation artifacts. Consequently, BTFs are much more demanding in terms of memory resources than SVBRDFs, so much so that they usually need to be compressed in order to fit within the graphic adapter's or even the computer's RAM. A tried way of achieving sufficient compression while retaining a high-quality BTF, which is also the most relevant to the present thesis, is based on *singular value decomposition* (SVD)

$$\mathbf{B} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^t$$

of the BTF in its matrix representation $\mathbf{B}$ (cf. Fig. 1.4). In this form, compression can be achieved by truncating $\mathbf{U}$, $\boldsymbol{\Sigma}$ and $\mathbf{V}$: let $k \in \mathbf{N}$, then by the Eckart-Young-Mirsky theorem [18], the rank-k truncated singular value decomposition

$$\mathbf{B} \approx \mathbf{U}_k \boldsymbol{\Sigma}_k \mathbf{V}_k^t$$

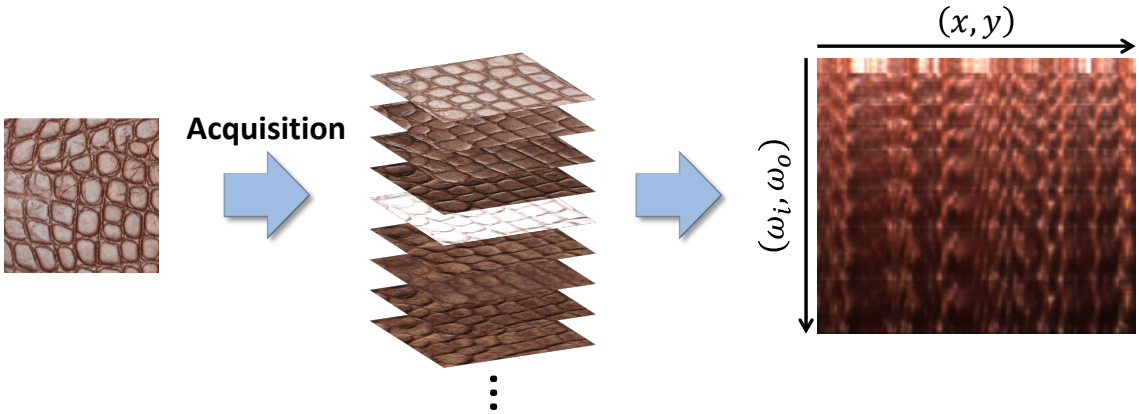


Figure 1.4: Left: A leather-like material sample. Center: A stack of bidirectional textures obtained by sampling the material's BTF. Right: Representation of the sampled BTF as a matrix.

minimizes

$$\operatorname{argmin}_{\mathbf{M}} \|\mathbf{B} - \mathbf{M}\|_{\mathrm{F}} \text{ s.t. } \operatorname{rk} \mathbf{M} = k,$$

where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm. The BTF is thus represented as a linear combination of *eigen-ABRDFs* $\mathbf{U}_k$, which can be considered a linear model for (precisely) the BTF at hand. Evaluation is still almost as straight-forward as for uncompressed BTFs, but comes with a performance penalty imposed by decompression. Note that typically some kind of non-linear transformation is applied to deal with the BTF data's dynamic range; cf. Ch. 3. Many other ways of modelling BTFs have been proposed, virtually always for the primary purpose of reducing storage requirements [113]. One of the present thesis' central points is that the above simple linear model serves particularly well as a foundation for more encompassing models suitable for all kinds of applications.

1.1.3 Acquisition of digital material appearance

Obtaining digital representations of a given real-world material's appearance can be arbitrarily convoluted, depending on a multitude of choices and compromises. Even though the type of reflectance function to be measured is arguably one of the most important choices to be made, the remaining ones still warrant vastly differing acquisition paradigms, depending on the compromises one is willing to accept. As a result, researchers and practitioners have come up with many types of acquisition devices and methods for the various types of reflectance functions as described above. Nonetheless, all of these devices and methods have in common that they rely on imaging systems, typically digital cameras, and artifical or natural lighting. The acquisition setup's complexity depends to some extent on the number of parameters of the reflectance function to be sampled: For instance, it may be possible to sample a 2D

texture with sufficient quality by taking a single image of the material, lit by an appropriate light source, whereas a 4D (or even 3D isotropic) BRDF may demand tens or hundreds of images of the material seen and lit from different angles to achieve a sufficient sampling rate, even when the material is painted on a spherical surface, allowing for a denser sampling of the specular highlight [24].

The case is not as clearly cut when it comes to SVBRDFs and BTFs, which have the same amount of parameters: Digital SVBRDFs are often represented by "textures" of parameters to analytical BRDF models, usually obtained from measured data using some kind of non-linear optimization, which can be achieved with a relatively low number of carefully selected samples [114]. Conversely, digital BTFs are almost exclusively represented in an image-based form, which may require a vast amount of samples in order to avoid interpolation artifacts. In both cases, the number of actual images to be taken by means of the imaging systems involved typically scales with the desired number of samples. Practically, this can lead to hundreds or thousands (in the case of SVBRDFs) and even tens of thousands (in the case of BTFs) of images to be required. Most state-of-the-art acquisition setups for SVBRDFs or BTFs thus contain multiple cameras in order to both reduce delays caused by moving a single camera or material sample to change the viewing angle and to allow for varying degrees of parallelization (but see below). Notable exceptions include a method for obtaining SVBRDFs from two pictures taken with a cellphone camera and its flash, and a BTF acquisition setup based on a kaleidoscope, allowing for multiple viewing angles in a single image ([42] and [21], respectively).

In the following I shall describe, in some generality, the process involved in acquiring digital representations of material BTFs as used to obtain the data for the research that culminated in the present thesis, along with the limitations it entails. In principle, it also applies to all the other types of reflectance functions described above; in practice, however, specialized methods will usually produce more satisfactory results. Relevant surveys (e.g. [60]) provide a broader and more in-depth introduction to the various acquisition paradigms.

Measuring BTFs

Common general-purpose BTF acquisition setups ("*camera domes*") consist of arrays of digital cameras and light sources distributed on a hemisphere above a sample holder in its center (cf. Fig. 1.5). Ideally, the setup is shielded from both environmental light in order to avoid unnecessary noise from subtracting the background, and from stray light caused by inter-reflections of the light emitted by the light sources. The setup needs to be calibrated both radiometrically and geometrically: Radiometric calibration entails determining the camera sensors' response curves, dark frames and color profile as well as the light sources' color profile and intensity distribution on the sample plane. Geometric calibration is used to determine the positions of the acquisition setup's relevant components with respect to a chosen coordinate system, such that for each captured image, the angles $\boldsymbol{\omega}_i$ and $\boldsymbol{\omega}_o$ of incoming and outgoing light, respectively, are known. The essential part of the physical acquisition process then consists of placing the sample to be measured on the sample holder

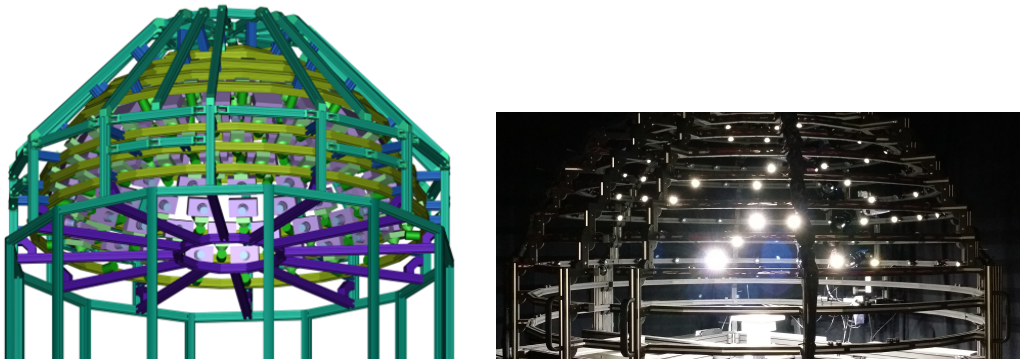


Figure 1.5: Left: *DOME I*, a highly parallel BTF acquisition device comprised of 151 consumer-grade digital cameras, the built-in flashes of which serve as light sources. Right: *DOME II*, a camera dome like DOME I, but with 11 industrial-grade digital cameras and 198 LEDs. The material sample is placed on a turntable which is rotated a user-definable number of times in order to achieve a sufficient sampling of viewing directions.

and taking one or usually several images of the material sample per combination of camera and light source, with all cameras taking images in parallel. Several images with different exposure times are needed in order to ensure that every part of the sample is correctly exposed in at least one image. When all desired images have been obtained, some further processing is required to bring the data into the form described in Sec. 1.1.2: Depending on the setup, an appropriate method such as de-mosaicing in the case of cameras with Bayer pattern sensors can be used to obtain RGB images. Subsequently, the per-$(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o)$ exposure series are combined into per-$(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o)$ HDR radiance maps [84]. Finally, after the radiance maps have been corrected radiometrically using the calibration data obtained beforehand, the trapezoid regions of interest within them are warped into rectangular shapes of a chosen size using the geometric calibration data which can then be stacked appropriately as to obtain tensor or matrix representations of the measured material's BTF. At this point, the digital BTF can be re-sampled for compression and/or rendering as described in Sec. 1.1.2 (cf. Fig. 1.4).

The two setups employed to obtain the data used in the present thesis (see Fig. 1.5 and below) are described in great detail in the publications that introduced them ([26] and [30], respectively).

Drawbacks & Limitations

While the process of obtaining a digital representation of a material's BTF as described above is straight-forward on an abstract level, it requires significant effort in practice. The time required for the post-processing step scales approximately linearly with the number of HDR radiance maps. The same holds for the data size of both the intermediate and the post-processed measured data, albeit with different constants. Worse, the actual measurement time scales approximately linearly not only with the number of light sources and images per

exposure series (assuming all cameras shoot simultaneously), but also with exposure time, which is dominated by the longest exposure time due to the way shutter times are usually chosen [29]. A dark, glossy material that exhibits self-shadowing may call for both many and long exposure times, leading to excessive time and storage requirements of possibly many hours, and terabytes of data [29]. One may seek to reduce these requirements simply be reducing the number of light sources or cameras and, by extension, of images captured and post-processed; however, this may come at the cost of severe undersampling of the material's BTF, leading to interpolation artifacts in rendering. Note that, conversely, there is no canonical "good" sampling rate – the more samples are used, the closer the digital BTF will reproduce the physical reality, albeit with diminishing returns at an application- and material-specific point.

Designing a BTF acquisition device entails, in fact, many choices like the above, all of which associated with some of the setup's attributes such as cost, size, power consumption, acquisition speed, output quality, etc. As an example, the quality of the acquisition setup's output depends on texture-space resolution, which in turn depends on parameters such as image sensor resolution, image noise, optical properties of the camera lenses, and the distance between cameras and material sample. All of these parameters also influence attributes such as cost and size.

The present thesis focuses on methods to reduce both measurement and total acquisition time, and to increase the texture-space resolution of the resulting digital BTFs. These attributes can be controlled via setup parameters: at any rate, measurement and total acquisition time depend on

- brightness of the light sources;
- photon efficiency of the cameras' sensors;
- size of the cameras' lenses' apertures;
- numbers of light sources and cameras;

texture-space resolution on

- size and pixel density of the cameras' sensors;
- size of the cameras' lenses' apertures;
- focal length of the cameras' lenses;
- distance between the cameras and the material sample.

Thus, both attributes can be controlled by varying the parameters they depend upon. However, the degree of control is bounded by both hard and soft constraints: for instance, for a given lighting setup, there are hard theoretical limits on sensor pixel size, depending on the signal-to-noise ratio one is willing to accept [115]. Once the limit is reached, in order to

reduce exposure times further, one may choose brighter light sources, which may or may not be available or affordable, or extend the size of the cameras' lenses' apertures, which may or may not be possible, or lead to reduced optical resolution or undesirably shallow depth of field. Overall acquisition time (including post-processing) can be decreased by reducing the number cameras and/or light sources and thereby of images to be captured and processed, at the cost of greater risk of noticeable interpolation artifacts in the output BTF. On the other hand, the greater their number, the bigger and more expensive the resulting setup, and the more difficult the digital BTFs become to deal with due to their sheer data size. But the distance between light sources and material sample and, by extension, the setup's size is also limited from below as to ensure approximately far-field illumination assumed for BTFs. If the cameras are placed much closer to the sample, they may cast shadows on its surface, which imposes similar constraints on their placement. The cameras' distance from the material sample, in turn, influences the texture-space resolution of the resulting BTFs. By choosing an appropriate focal length, a desired degree of resolution may be achieved; however, the larger the focal length, the smaller the part of the material sample seen by the cameras, and the narrower the cameras' aperture, leading to longer exposure times. In fact, there is a somewhat delicate balance between these parameters, particularly with respect to texture-space resolution [116].

To a certain degree, practical BTF acquisition setups of the type considered here therefore always suffer one way or another from the particular compromises made during their conception.

1.2 Contributions & Outline

The aim of the present thesis is to alleviate the above limitations inherent to the process of measuring BTFs (and, to a lesser degree, SVBRDFs, which can be obtained from measured BTFs). All of the present thesis' contributions presented in the following use data-driven appearance models one way or another in order to improve upon and enhance state-of-the-art material appearance acquisition devices through computational means.

The present thesis is a cumulative thesis; that is, its content is comprised mostly of a number of peer-reviewed publications to which I was the primary contributor. Specifically, the following articles are summarized in the subsequent chapters:

[1] D. den Brok, H. C. Steinhausen, M. B. Hullin, and R. Klein, *Patch-Based Sparse Reconstruction of Material BTFs*, Journal of WSCG 22 (2014) 83,

[2] D. den Brok, H. C. Steinhausen, M. B. Hullin, and R. Klein, "Multiplexed Acqusition of Bidirectional Texture Functions for Materials,", *Measuring, Modeling, and Reproducing Material Appearance II (SPIE 9398)*, vol. 9398, doi: 10.1117/12.2078396, [1]

[1] best student paper award

[3] D. den Brok, H. C. Steinhausen, and R. Klein, "Fast multiplexed acquisition of high-dynamic-range material appearance," *Vision, Modeling & Visualization*, 2015 151, doi: 10.2312/vmv.20151270,

[4] D. den Brok, M. Weinmann, and R. Klein, *Rapid material capture through sparse and multiplexed measurements*, Computers & Graphics 73 (2018) 26, doi: 10.1016/j.cag.2018.03.003,

[5] D. den Brok, S. Merzbach, M. Weinmann, and R. Klein, "Per-Image Super-Resolution for Material BTFs," *2020 IEEE International Conference on Computational Photography (ICCP)*, 2020 1, doi: 10.1109/ICCP48838.2020.9105256.

Note that [2] did not undergo full peer review, but it is included because it serves as a preface to [3], which extends it. I contributed to varying degrees to the following publications as well, but either I was not the primary contributor, or they were intended as summaries of previous or preliminary work to be presented to a workshop audience, hence they are not included in the present thesis:

[6] H. C. Steinhausen, D. den Brok, M. B. Hullin, and R. Klein, *Acquiring Bidirectional Texture Functions for Large-Scale Material Samples*, Journal of WSCG 22 (2014) 73,

[7] H. C. Steinhausen, D. den Brok, M. B. Hullin, and R. Klein, "Extrapolation of bidirectional texture functions using texture synthesis guided by photometric normals," *Measuring, Modeling, and Reproducing Material Appearance 2015*, vol. 9398, SPIE, 2015 81,

[8] D. den Brok, M. Weinmann, and R. Klein, "Linear Models for Material BTFs," *Eurographics Workshop on Material Appearance Modeling*, The Eurographics Assocation, 2015 15,

[9] H. C. Steinhausen, D. den Brok, M. B. Hullin, and R. Klein, "Extrapolating Large-Scale Material BTFs under Cross-Device Constraints," *Vision, Modeling & Visualization*, The Eurographics Assocation, 2015,

[10] M. Weinmann, D. den Brok, S. Krumpen, and R. Klein, "Appearance Capture and Modeling," SIGGRAPH Asia 2015 Courses, ACM, 2015,

[11] D. den Brok, M. Weinmann, and R. Klein, "Towards Sparse and Multiplexed Acquisition of Material BTFs," *Workshop on Material Appearance Modeling*, The Eurographics Assocation, 2017,

[12] H. C. Steinhausen, D. den Brok, S. Merzbach, M. Weinmann, and R. Klein, "Data-driven Enhancement of SVBRDF Reflectance Data", Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Vol. 1: GRAPP, 2018,

[13] C. Callenberg, A. Lyons, D. den Brok, R. Henderson, M. B. Hullin, and D. Faccio, "EMCCD-SPAD Camera data fusion for high spatial resolution time-of-flight imaging," *Imaging and Applied Optics 2019*, Optical Society of America, 2019,

[14] C. Callenberg, A. Lyons, D. den Brok, A. Fatima, A. Turpin, V. Zickus, L. Machesky, J. Whitelaw, D. Faccio, and M. B. Hullin, *Super-resolution time-resolved imaging using computational sensor fusion*, Scientific Reports 11 (2021) 1689.

CHAPTER 2

Patch-based sparse reconstruction of material BTFs

The research summarized in this chapter has been published as

> D. den Brok, H. C. Steinhausen, M. B. Hullin, and R. Klein, *Patch-Based Sparse Reconstruction of Material BTFs*, Journal of WSCG 22 (2014) 83.

For consistency with the subsequent chapters, I have slightly modified the notation in the following summary and use the symbol $\mathbf{S}$ for the measurement matrix instead of $\mathbf{M}$ as it occurs in the publication.

2.1 Summary

Arguably two of the biggest drawbacks of digital image-based material representations such as BTFs are the time and storage capacity it takes to acquire them from given real-world material samples. In both cases, the number n of images that have to be captured in order to avoid obvious interpolation artifacts in renderings using the resulting digital BTF is the dominant factor. However, all these images depict the self-same material sample, albeit as seen and lit from varying angles. It is hence reasonable to assume that a given per-$(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o)$-texture can at least approximately be inferred from others, for instance using an appropriate model to describe the measured BTFs. Indeed, as outlined in Sec. 1.1.2 there is a straight-forward data-driven linear model given by the truncated singular value decomposition (SVD) $\mathbf{B} \approx \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k$, where $\mathbf{B}$ is the measured BTF represented as a matrix. When only $n_s \ll n$ images of the material sample are taken, $\mathbf{B}$ and, by extension, $\mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^t$ are unknown in their entirety. Instead, what is measured can be described in terms of $\mathbf{B}$ and a *measurement matrix* $\mathbf{S} \in \{0,1\}^{n_s \times n} \subset \mathbf{R}^{n_s \times n}$ with $\mathbf{S}\mathbf{S}^t = \mathbf{1}$. The goal is then to recover an approximation of $\mathbf{B}$ from the actual measurement $\mathbf{S}\mathbf{B}$. Given a linear model $\mathbf{U}$, this can be achieved by solving $\mathbf{S}\mathbf{U}\mathbf{V}^t = \mathbf{S}\mathbf{B}$ for $\mathbf{V}$; then simply $\mathbf{B} \approx \mathbf{U}\mathbf{V}^t$.

This approach raises a number of questions:

- Apart from $\mathbf{U}_k$, which is unknown *a priori*, what is a good model $\mathbf{U}$ for this scenario?
- Given a model $\mathbf{U}$, what is a good choice of $\mathbf{V}$?
- Which and how many per-$(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o)$ yield pleasing reconstruction results, i.e. what is a good choice of n_s and $\mathbf{S}$?

In the publication summarized in this section [1], we give practical answers to these questions.

The model $\mathbf{U}_k$ of the first k eigen-ABRDFs of $\mathbf{B}$ may be unknown beforehand because so is $\mathbf{B}$ itself, but when some or many BTFs $\mathbf{B}_1, \ldots, \mathbf{B}_m$ of other materials have been measured in their entirety before, a linear model $\mathbf{U}$ can be derived from these, for example by computing a truncated SVD for all $\mathbf{B}_i$ at once, or separately and merging them thereafter. We demonstrate that a model $\mathbf{U}$ obtained this way indeed generalizes to BTFs $\mathbf{B} \notin \{\mathbf{B}_i\}$, albeit at the cost of a greater but tolerable number of model parameters required.

Unfortunately, the greater the number k of model parameters, the more samples are usually needed to fit them robustly to a measurement, or else the resulting system $\mathbf{SUV}^t = \mathbf{SB}$ of linear equations may become close to being under-determined or even beyond. This can often be mitigated by means of additional regularization terms, but we chose a different approach and instead modified the model $\mathbf{U}$. By computing $\mathbf{U}$ not on a set of ABRDFs from different BTFs, but on a set of small quadratic texture-space *patches* of ABRDFs, we both incorporate some knowledge about how different spatial structures such as height differences translate to effects in ABRDFs, and drastically increase the number of equations in the linear system to be solved. That way, we are able to determine $\mathbf{V}$ much more robustly even for small numbers n_s of samples, simply via

$$\mathbf{V}^t_{\text{patch}} = \left(\mathbf{S}_{\text{patch}}\mathbf{U}_{\text{patch}}\right)^{\dagger} \mathbf{S}_{\text{patch}}\mathbf{B}_{\text{patch}}$$

where p is the patch size, $\mathbf{S}_{\text{patch}} = \mathbf{1}_{p^2} \otimes \mathbf{S}$ the Kronecker product of the $p^2 \times p^2$ identity matrix and $\mathbf{S}$, and $(\mathbf{S}_{\text{patch}}\mathbf{U}_{\text{patch}})^{\dagger}$ the Moore-Penrose pseudo-inverse of $\mathbf{S}_{\text{patch}}\mathbf{U}_{\text{patch}}$.

It remains to determine $\mathbf{S}$ and n_s. For a fixed value of n_s, we follow Matusik et al. [23] and minimize the condition number $\kappa(\mathbf{S}_{\text{patch}}\mathbf{U}_{\text{patch}})$ by iteratively replacing single lines of $\mathbf{S}$ until convergence to a local minimum. That way, only the n_s pairs $(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o)$ most conducive to solving the system of linear equations robustly are sampled.

In our experiments, we found that using the proposed method we were able to produce visually pleasing results with as little as 6 % of the original samples. Note that depending on the degree of parallelism in the acquisition setup, this does not necessarily mean that the measurement only takes 6 % of the time of a full measurement. For fully parallel setups, we provide a heuristic sampling strategy with the same number of samples, but chosen such that one and the same subset of the available cameras is used throughout to exploit parallelism, at the cost of a slight reduction in reconstruction quality.

2.2 Author's contributions

My own contributions to the above research encompass experimentation with different models, ways to obtain them efficiently despite the amount of data to be processed, various regularization methods, implementing most of the algorithms involved, evaluating the results and, finally, summarizing the findings in the resulting publication [1]. My co-authors provided guidance, helpful discussions, and writing assistance.

CHAPTER 3

Fast multiplexed acquisition of high-dynamic-range material appearance

The research summarized in this chapter has been published as

> D. den Brok, H. C. Steinhausen, M. B. Hullin, and R. Klein, "Multiplexed Acqusition of Bidirectional Texture Functions for Materials,", *Measuring, Modeling, and Reproducing Material Appearance II (SPIE 9398)*, vol. 9398, doi: 10.1117/12.2078396,

and

> D. den Brok, H. C. Steinhausen, and R. Klein, "Fast multiplexed acquisition of high-dynamic-range material appearance," *Vision, Modeling & Visualization*, 2015 151, doi: 10.2312/vmv.20151270.

3.1 Summary

When measuring dark materials, or materials exhibiting a large amount of self-shadowing, it is crucial that as much of the reflected light as possible reaches the cameras' sensors in order to keep shutter times low. However, some of the compromises involved in devising BTF acquisition setups, described in more detail above (cf. Sec. 1.1.3), may lead to very little light actually being sensed by the cameras, for instance when small apertures are used to ensure a sufficiently large depth of field. Correct exposure under all circumstances can then only be achieved through exposure series including very long shutter times, possibly on the order of minutes [29]. In conjunction with the huge number of images to be captured in the first place, the resulting overall measurement times quickly become excessive. In a fixed acquisition setup bound by the above constraints, there is thus little choice but to light the material sample using several light sources from different angles at once. That way, both more light arrives at the material surface and is reflected towards the cameras, and regions shadowed when the

scene is lit by one particular light source have a chance of being lit by at least one other light source. In fact, when lit by *all* light sources relatively densely covering the hemisphere above the material sample at once, the dynamic range of the resulting images may be reduced to such an extent that a single, short exposure is sufficient for all images. However, there is no obvious way of deriving the desired images of the material lit by single light sources from images obtained this way. The case is different when consecutive, changing illumination *patterns* of several (but not all) light sources at once are used; an approach called *multiplexed* measuring.

Indeed, let $n_i \in \mathbf{N}$ be the number of light sources. In analogy to the research presented in Ch. 2, an illumination-multiplexed BTF measurement $\mathbf{B}_{\text{multiplexed}}$ can be modelled in terms of a measurement matrix $\mathbf{M} \in \{0,1\}^{n_p \times n_i} \subset \mathbf{R}^{n_p \times n_i}$, with $n_p \in \mathbf{N}$ the number of patterns to be used:

$$\mathbf{B}_{\text{multiplexed}} = \mathbf{MB},$$

with $\mathbf{B}$ and $\mathbf{B}_{\text{multiplexed}}$ re-shaped appropriately. If $\mathbf{M}$ is full-rank, $\mathbf{B}$ can be solved for exactly ($n_p = n_i$) or at least approximately ($n_p \neq n_i$). In the above example (assuming $n_p = n_i$) where all lights are used simultaneously, $\mathbf{M}$ is singular and therefore unsuitable. An obvious modification which makes $\mathbf{M}$ invertible is $\mathbf{M} := \mathbf{M} - \mathbf{1}_{n_i}$, corresponding to patterns of all light sources but a single one varying from pattern to pattern. It turns out that the reconstructions obtained from real-world images using these patterns typically have a very low signal-to-noise ratio (SNR). The reason is that physical measurements are inherently subject to noise, some of which may be signal-independent, some of which may not. It can be shown that in the case of typical signal-independent noise, the ratio of root mean square (r.m.s.) SNRs of non-multiplexed measurement and of reconstruction from multiplexed measurements ("de-multiplexing") can be computed as

$$\sqrt{\frac{\text{MSE}}{\text{signal}} \frac{\text{signal}}{\text{MSE}_{\text{de-multiplexed}}}} = \sqrt{\frac{\text{MSE}}{\text{MSE}_{\text{de-multiplexed}}}} = \sqrt{\frac{n_i}{\text{tr}\,[M^t M]^{-1}}},$$

where MSE and $\text{MSE}_{\text{de-multiplexed}}$ are the corresponding mean square errors [33]. For the above choice of $\mathbf{M}$, this actually amounts to a slight numerical improvement of SNR in the reconstruction, with a ratio very close to 1. However, in imaging setups, signal-dependent *photon noise* is inevitable. Photon noise follows a Poisson distribution and therefore increases with the square root of the overall amount of light. For the patterns suggested above, this amounts to a factor of (approximately) n_i in the denominator in the left-hand side of the above equation and thus a *decrease* of SNR by a factor of about $\frac{1}{\sqrt{n_i}}$, which leads to overwhelming amounts of noise in the reconstructions for practical values of n_i.

In the publications summarized in this section [2, 3], we seek to leverage the general multiplexing approach nonetheless and aim at answering the following questions:

- Are there patterns which provide a reasonable balance between noise in the reconstructions and acquisition speed?

- Can reconstruction noise additionally be reduced further, e.g. (additionally) by means of a linear model similar to the one presented in Ch. 2?

- If so, is there a way to reduce the time required to actually obtain a suitable model in the first place?

Indeed, general theory on multiplexed measuring states that multiplexing patterns derived from *Hadamard matrices* are at least close to optimal with respect to the improvement of signal-to-signal-independent-noise ratio. Unfortunately, this advantage is counteracted by the degradation of signal-to-signal-dependent-noise ratio, which can be quantified as $\frac{1}{\sqrt{2}}$ in the worst case [33]. However, the resulting patterns consist of $\frac{n_i}{2}$ active light sources spread over the hemisphere, usually providing greatly increased acquisition speed. As a measure to reduce the amount of noise in the reconstructions, we find fitting the reconstructed BTFs to a linear model $\mathbf{U}$ similar to the one described in the previous section to be effective; i.e.

$$\mathbf{B}_{\text{denoised}} = \mathbf{U}\left(\mathbf{U}^{\dagger}\mathbf{B}\right).$$

Additionally, the effectiveness of this approach can be enhanced by determining reconstructed per-(ω_i, ω_o) textures with particularly high variance, which are likely to be too corrupted by noise, and casting the above fitting problem as a "sparse" reconstruction problem with the high-variance textures as missing values; i.e.

$$\mathbf{B}_{\text{denoised}} = \mathbf{U}\left((\mathbf{SU})^{\dagger}\mathbf{SB}\right).$$

Using these approaches, we find Hadamard-derived illumination patterns to provide a reasonable balance between reconstruction quality and acquisition speed. In our experiments, we typically only need a single exposure step of under 100 ms instead of four of up to tens of seconds, reducing raw measurement times by up to 95 %.

Obtaining a BTF database suitable for computing a linear model $\mathbf{U}$ may be an obstacle in its own right. For the acquisition setup employed here, and described in Fig. 1.5 and the relevant literature [30], which captures images in parallel only for a small slice of the hemisphere of viewing angles, it is easy to determine beforehand which images will be particularly bright due to the Fresnel effect and therefore cause the most signal-dependent noise. In order to acquire a high-quality database using the proposed method, acquisition speed can be traded for higher-quality BTFs by sub-dividing the illumination hemisphere into sectors of (roughly) equal size and capturing images in the sector affected by the Fresnel effect without multiplexing. The remaining sectors can then be measured using multiplexed illumination with smaller n_i and thus lower amounts of noise at the cost of longer shutter times. We find that this approach still reduces raw measurement time by about 60 %, and that the resulting linear model $\mathbf{U}'$ is similarly effective for the purpose of noise reduction.

3.2 Author's contributions

My contributions to the research described in this chapter consist of re-writing parts of the calibration and post-processing pipelines to accomodate multiplexed illumination, experimentation with illumination patterns and de-noising algorithms, acquisition of multiplexed and non-multiplexed BTF databases, evaluating the reconstruction results and, finally, summarizing the findings in the resulting publications [2, 3]. My co-authors provided guidance, helpful discussions, and writing assistance.

CHAPTER 4

Rapid material capture through sparse and multiplexed measurements

The research summarized in this chapter has been published as

> D. den Brok, M. Weinmann, and R. Klein, *Rapid material capture through sparse and multiplexed measurements*, Computers & Graphics 73 (2018) 26, doi: 10.1016/j.cag.2018.03.003.

4.1 Summary

In the preceding two chapters, I describe two methods that can help accelerate the cumbersome process of acquiring BTFs based on reducing the amount of HDR radiance maps of the material sample to be taken, and the number and duration of exposure steps required in order to ensure correct exposure. At first glance, it seems that these methods are somewhat orthogonal and can, therefore, be easily combined, as in

$$\mathbf{B} \approx \mathbf{U} \left[(\mathbf{SMU})^{\dagger} (\mathbf{SMB}) \right],$$

with $\mathbf{U}$, $\mathbf{S}$ and $\mathbf{M}$ as in the previous chapter. This corresponds to measurements both illumination-multiplexed and sparse with respect to the light sources and cameras available in the given acquisition setup. However, there is a severe obstruction: both methods require a linear model $\mathbf{U}$ derived from a database of measured BTFs with possibly wide dynamic range. Computing $\mathbf{U}$ by means of a truncated SVD on such data is known to lead to visible artifacts due to an over-emphasis of specular highlights [23, 50]. An established counter-measure is to apply some kind of non-linear transformation to the data beforehand that reduces the dynamic range, commonly the logarithm:

$$\log \mathbf{D} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{t},$$

and in fact this transformation was used in the methods introduced in the previous chapters, too. One can thus only reasonably expect $\log \mathbf{B}$, not $\mathbf{B}$ itself, to be adequately representable in terms of $\mathbf{U}$. But $\log(\mathbf{SMB}) = \mathbf{S}\log(\mathbf{MB}) \neq \mathbf{SM}\log\mathbf{B}$ in general, so at any rate, the latter can only be determined if $\mathbf{S}$ is chosen to incorporate all n_p illumination patterns (because then $\mathbf{SMB}$ can be de-multiplexed), which defeats the purpose of sparse acquisition.

In the published research summarized in this section [4], we seek to answer the following questions:

- Is there a model $\mathbf{U}$ with similarly favorable properties as the one from Ch. 3, but which allows for simultaneous de-multiplexing and sparse reconstruction as described above?
- If so, how can it be computed efficiently?

A straight-forward approach to the problem is to obtain $\mathbf{U}$ directly from $\mathbf{D}$. Then the terms in the above equation for a sparse, multiplexed measurement become "compatible" and may thus yield satisfactory results. It remains to determine a way of dealing with the data's wide dynamic range. We propose to replace the commonly used optimization problem

$$\mathbf{U},\ \mathbf{V} = \operatorname{argmin}_{\tilde{\mathbf{U}},\tilde{\mathbf{V}}} \| \log \mathbf{D} - \tilde{\mathbf{U}}\tilde{\mathbf{V}} \|_{\mathrm{F}}$$

with

$$\mathbf{U},\ \mathbf{V} = \operatorname{argmin}_{\tilde{\mathbf{U}},\tilde{\mathbf{V}}} \| \mathbf{W} \odot (\tilde{\mathbf{U}}\tilde{\mathbf{V}} - \mathbf{D}) \|_{\mathrm{F}},$$

where $\odot$ is the element-wise matrix product, and $\mathbf{W}$ a matrix of *weights*, and in both cases the minimization parameters $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ have k columns and rows, respectively. This effectively allows for fine-tuning the metric used when computing $\mathbf{U}$. In particular, if $\mathbf{W}$ is chosen as $\mathbf{W} := \mathbf{D}^{-1}$, this amounts to minimizing the per-element *relative* L2 error $\frac{|x-\tilde{x}|}{|x|}$. A metric like this ensures that particularly bright parts of the BTF, like highlights or grazing angles with strong Fresnel effect, are not overly emphasized in the fitting process.

However, contrary to the standard least-squares problem, there is no canonical way of solving the resulting optimization problem. We propose an *alternating least-squares* approach, alternatingly solving for $\mathbf{U}$ and $\mathbf{V}$ as vectors until convergence to a (typically local) minimum. The resulting linear systems of equations are sparse, but impractical to handle nonetheless. Fortunately, their sparsity structures allow for solving on a per-$(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o)$ and per-texel basis, respectively, which reduces the complexity such that the partial problems can be handled easily. As with the SVD-based approach, this is most easily achieved on a per-material basis. $\mathbf{U}$ can then be obtained by computing an SVD of the concatenated per-material $\mathbf{U}^{(i)}$.

Finally, as mentioned in Ch. 2, reconstruction from sparse measurements works best with an appropriate regularization method. Following Nielsen et al. [54], who introduced it to sparse reconstruction of BRDFs, we employ a *Tihonov* regularization term:

$$\mathbf{V} = \operatorname{argmin}_{\tilde{\mathbf{V}}} \|\mathbf{SMU}\tilde{\mathbf{V}} - \mathbf{SMB}\|_{\mathrm{F}} + \|\boldsymbol{\Gamma}\mathbf{V}\|_{\mathrm{F}}$$

which I had experimented with with some success in the context of BTFs prior to the publication of the referenced work, but which had not made it into a publication of my

own up until that point. $\mathbf{\Gamma}$ can be chosen arbitrarily; similar to Nielsen et al., we choose $\mathbf{\Gamma} = \mathbf{\Sigma}^{-1}$, where $\mathbf{\Sigma}$ is the diagonal matrix of singular values obtained along with $\mathbf{U}$. This choice penalizes large deviations from the database's singular value distribution and hence ensures a plausibel BTF.

In our experiments, we found our alternating least-squares algorithm to converge quickly and to local minima reasonable in the sense that they always adequately represented the BTFs from the validation set, albeit at the cost of an increased number of parameters for approximately the same quality when compared to the conventional models. When put into practice, we are able to produce visually pleasing BTFs from sparse and multiplexed measurements taken in only 15–30 minutes, where the "brute-force" measurements take as much as 10–23 hours, and separately sparse and multiplexed measurements 1–4 hours.

4.2 Author's contributions

My contributions to the above research entail development, implementation and fine-tuning of the involved algorithms, evaluation of the reconstruction results, and writing most parts of the actual publication. My co-authors provided guidance, helpful discussions, literature research, and writing assistance, particularly the section on related work.

CHAPTER 5

Per-image super-resolution for material BTFs

The research summarized in this chapter has been published as

> D. den Brok, S. Merzbach, M. Weinmann, and R. Klein, "Per-Image Super-Resolution for Material BTFs," *2020 IEEE International Conference on Computational Photography (ICCP)*, 2020 1, doi: 10.1109/ICCP48838.2020.9105256.

5.1 Summary

In the previous chapters, measurement matrices were applied from the left to BTF data represented as matrices with columns corresponding to ABRDFs. Mathematically, it is admissible to apply matrices from the right and thus to the BTF's bidirectional textures, and in fact this operation may have a perfectly valid physical interpretation depending on the choice of measurement matrix $\mathbf{M}$. For instance, $\mathbf{M}$ can be chosen such that it corresponds to convolution with a camera's point spread function and subsequent down-sampling. Solving the resulting linear system amounts to super-resolution, i.e. up-sampling and de-blurring, for single images. In the case of BTFs, this corresponds to texture-space super-resolution performed on each texture separately, which has been shown to work reasonably well even on compressed representations [68]. However, while super-resolution is an important technique which helps strike a more favorable balance of parameters in imaging systems such as BTF measurement devices (cf. Sec. 1.1.3), it also is a hard problem to solve, because it is under-determined by nature. Consequently, although it has been studied in general for decades, the problem both in general and in special cases of interest is still a topic of ongoing research. Following the general trend, many single-image super-resolution algorithms today make use of modern "deep learning" techniques such as neural networks to great effect. For several reasons, it is not clear *a priori* that this approach also lends itself to per-bidirectional texture super-resolution for BTFs:

Firstly, the resulting super-resolved images are often convincing in that they are plausible and more detailed than what "classical" algorithms produce, but they need not always resemble the ground-truth very closely. A BTF's per-$(\boldsymbol{\omega}_i, \boldsymbol{\omega}_o)$ textures vary in a physically meaningful way, and it is unclear whether they still do when super-resolved by a "hallucinating" single-image algorithm, applied to each texture separately. If they don't, the resulting super-resolved BTF may exhibit annoying artifacts when rendered.

Secondly, BTFs typically cover a wider dynamic range than the images usually considered in single-image super-resolution. Whether the same algorithms carry over to the HDR case needs to be determined.

Finally, and on the other hand, BTF data is much less diverse than a completely random sample of real-world images. In both this and the previous case, it is reasonable to assume that a neural network trained on BTF data alone outperforms the same network trained on real-world images. However, it may be difficult to find appropriate ground-truth data: in a fixed acquisition setup, the only available data may be data produced by the setup itself. Down-sampling that data in order to produce training data only tells the network about the transition from the down-sampled resolution to the setup's actual resolution, but the intention is to exceed that resolution.

Our aim in the research project summarized in the present chapter is to investigate whether "deep learning"-based single-image super-resolution algorithms can be leveraged to full BTFs. In particular, we provide initial answers to the following questions:

- Are typical single-image super-resolution networks suitable for the high dynamic range (HDR) radiance maps BTF acquisition devices produce?
- Is ground-truth data with a resolution exceeding the setup's required, or is the problem scale-invariant to some extent?
- Does the image-by-image approach lead to visible artifacts in the resulting BTFs?

To this end, we construct a convolutional neural network by enhancing a classical super-resolution network with current best practices. We train and test the network on a database of HDR radiance maps down-sampled once and twice. From the resulting high-resolution test images, we produce high-resolution BTFs which we evaluate against both the ground-truth and a state-of-the-art BTF super-resolution algorithm, and with respect to the above questions. We find that we can answer the questions in favor of the chosen algorithm, and that our algorithm outperforms the state-of-the-art.

5.2 Author's contributions

My contributions to the above research include development, implementation and fine-tuning of both the algorithm itself and visualization of the results, experimentation with different models, evaluating the results and summarizing them in the corresponding publication. My co-authors Michael Weinmann and Sebastian Merzbach both suggested various possible

improvements to the CNN, some of which turned out to be critical and were thus used in the final version of the algorithm. Sebastian Merzbach also helped with debugging the algorithm's implementation and writing and debugging the visualization code used for evaluation and presentation. Michael Weinmann did a large amount of literature research and helped with writing the "Related Work" section. All co-authors furthermore provided guidance and helpful discussions.

CHAPTER 6

Conclusion

The publications corresponding to the above summaries each contain an overview of the relevant related work at the time they were written. In the meantime, a number of publications have appeared which are immediately relevant in the context of the present thesis, some of which I have picked as representatives of lines of similar work.

6.1 Recent developments

In research conducted in our group and with my participation, Steinhausen et al. [12] use a linear BTF model similar to ours described above (cf. Ch. 2) to obtain BTFs with high angular resolution from a commercially available low-angular-resolution SVBRDF acquisition device. The advantage is that the acquisition device is readily available on the market and faster and much more compact than high-angular-resolution acquisition devices specifically for BTFs.

As an interesting and purely optical approach to increasing the texture-space resolution of BTF acquisition setups, Havran et al. [116] investigate Scheimpflug optics. They observe that most common setups are, in fact, limited in optical resolution by diffraction occuring due to the narrow camera apertures required to keep the entire material sample in focus at grazing viewing angles. By using Scheimpflug optics, the focal plane can be tilted in order to mitigate this problem. They further show that, using an additional anamorphic attachment, the loss of resolution at the "far" end of the material sample resulting from the perspective transformation can be counteracted. By combining both techniques, they show that the system's optical resolution can be enhanced by a factor of as large as approximately 2.

Our choice of illumination patterns for illumination-multiplexed BTF acquisition (cf. Ch. 3) is somewhat arbitrary, albeit effective. Kang et al. ([117]) propose a method for learning illumination patterns for the purpose of efficiently acquiring shape and reflectance of 3D objects endowed with an SVBRDF-like surface material. The approach is based on a deep neural network with a linear layer corresponding directly to the acquisition setup's light sources. Consequently, the resulting patterns are not binary and hence do not apply directly to typical BTF acquisition setups. It should be possible to either adapt the algorithm, or to

construct a BTF acquisition setup with dimmable light sources.

Following their 2019 article on the compression and interpolation of a single BTF using neural networks [83], and in analogy to our own work on linear models for BTFs [1], Rainer et al. demonstrate that a similar model trained on database of BTFs with varying angular samplings generalizes to a wide range of BTFs of distinct materials at once, regardless of their angular samplings [118]. Sampling-invariance is achieved by adding a Multi-Layer Perceptron as a pre-processing step. The authors claim that their model lends itself to applications such as BTF synthesis in the sense of extrapolation along the spatial domain and sparse acquisition, and provide initial experimental evidence to support their claim. However, they note that lack of training data is one of the major obstacles to achieving even better results.

Another very promising neural BTF model has been proposed by Kuznetsov et al. [119] Contrary to all prior BTF models, it is capable of incorporating geometric effects most prominent at grazing viewing angles, such as parallax effects and silhouette. As a result, objects rendered with such a "Silhouette BTF" (SBTF) as surface material look much more realistic at the borders than traditional BTFs in the absence of advanced and computationally expensive rendering techniques. The research's focus lies on the model's performance itself, both in terms of reconstruction quality and practicability in rendering applications; whether it also boasts similar qualities as the model introduced by Rainer et al. [83] warrants further investigation.

Apart from these, the field of material appearance in general has seen further developments not immediately relevant in the present context, including modelling of material representations other than BTFs using neural networks, and digital material appearance acquisition techniques in general. I thus refer the reader to corresponding recent surveys such as the one by Dong et al. [111] The research described in Ch. 5 strongly depends on single-image super-resolution; however, the model itself was chosen not with the aim of picking the currently "best" one in whatever terms, but to demonstrate the approach's feasibility. I therefore consider enumerating the large, ever-growing body of additional work on this topic of great general interest out of the present thesis' scope and refer the reader to recent surveys, e.g. the ones by Yao et al. [120] and Bashir et al. [121]

6.2 Discussion & Outlook

In the present thesis, I have introduced a number of algorithms and techniques which lay a foundation for significant improvements to the process of digitally capturing material appearance. Arguably all of these were pioneering works, not necessarily in the sense of "ground-breaking", but in that they brought qualitative change to the field of BTF acquisition as opposed to mere quantitative improvements: prior to them, to the best of my knowledge, a model expressive enough to represent a wide range of BTFs at once, sparse acquisition (cf. Ch. 2), multiplexed acquisition (cf. Ch. 3), the combination of the latter two (cf. Ch. 4), and texture-space super-resolution based on neural networks (cf. Ch. 5) had not been considered in the context of BTFs. Consequently, the goal of the works ultimately was not to provide the optimal BTF acquisition strategy, whatever that might be, but to demonstrate that certain ideas are both feasible and effective within this context in the first place. This naturally leaves room for further exploration and improvements.

The most apparent approach is to take the techniques as they are from research into development (as in "research and development", "R&D") in order to optimize them for practical use and make them more robust. All of our methods and algorithms are subject to a number of parameters that can be tuned such that an optimal balance between parameters, subject to constraints imposed by practical concerns, is struck, similar to the technical parameters that have to be considered when designing an acquisition device (cf. Sec. 1.1.3). As a (simplified) example, our linear model has itself a certain number of model parameters that need to be fitted, and the greater their number, the better the model represents BTFs in general, but the more samples are required in order for the fitting process to be robust, and their actual number is, in turn, a parameter itself. Similarly, the greater the number of LEDs in the acquisition setup, the greater the reduction of shutter times achievable through multiplexing, but the greater the amount of noise to be dealt with.

Optimized as in the above or not, our algorithms all to some degree enable the relaxation of constraints imposed upon acquisition setup design. By extension, they pave the way to more efficient acquisition setups which yield digital material BTFs of a quality similar to what is possible with a "comprehensive" setup used to obtain the models underlying our algorithms. Such a setup might be designed using only a small number of light sources and/or cameras, allowing for more compact, light-weight, and possibly more cost-effective devices. In fact, gonioreflectometers seem feasible for BTF acquisition again, given that devices with a low degree of parallelization benefit the most from our sparse acquisition method. Additionally, we already demonstrated that a commercially available industrial-grade SVBRDF acquisition device small enough to fit on a large desk can be used to measure a material's BTF using one of our methods [12]. Conversely, one might actually be interested in building larger devices or use camera lenses with a wider field of view than one would usually consider: it is a well-known problem – which we have also dealt with in separate publications [6, 9] – that most acquisition devices are limited to material patches of a spatial extent that may be too small in certain practical use cases. Our super-resolution algorithm can be used to compensate for the loss of texture-space resolution that comes with the above solutions to

this problem and thus allows for much larger material samples to be measured.

On the algorithmic side, there is a lot of room for further experimentation: When it comes to acquisition setups, it might be possible to devise an *adaptive* sparse sampling strategy which stops when, e.g., a certain quality criterion is met, such that fewer samples may be needed in many cases. While illumination patterns based on Hadamard matrices have been shown to work well, there may be choices which strike a better balance between amount of noise after reconstruction and reduction of shutter times, possibly adaptive as well. Moreover, linear models are among the most simple ways of modelling BTFs, but even though more sophisticated models may seem intriguing, linear models may have advantages that we have not investigated thoroughly yet. For instance, having a "global" model for a wide range of BTFs may facilitate defining a meaningful (pseudo-)metric on the space of BTFs, making it easier to compare BTFs, or search for related BTFs in a material database, or interpolate between (more than two) BTFs, similar to what Ruiters et al. have demonstrated [122]. Apart from the type of model used, there are a number of parameters which can possibly be improved upon, such as the metric when used to fit the model, which in turn depends on color space, reduction of dynamic range, and so on; or the regularization terms. Note that the optimization problem that occurs in our algorithm combining sparse and multiplexed acquisition contains a weight matrix $\mathbf{W}$ which can be easily modified to accomodate more sophisticated metrics. Our super-resolution algorithm, in turn, is inherently also a compression technique, given that only at most $\frac{1}{4}$ of the data is necessary in order to obtain visually pleasing BTFs of a certain resolution. While it can be used as is for this particular purpose, it currently cannot replace compression methods such as the truncated SVD described in Sec. 1.1.2 both in terms of compression factor and suitability for real-time purposes. It might be possible to combine both techniques and achieve reasonable decompression performance at least suitable for offline rendering.

Again on the algorithmic side, there might still be considerable untapped potential in (generalizations of) the optimization problem introduced in Ch. 4. Apart from the rather technical suggestions in the previous paragraph, I strongly believe it is possible to expand the idea sketched in Ch. 5 in order to achieve multi-view super-resolution for BTFs. The key observation is that a "measurement matrix" can be also used to model the process of rendering an assumed underlying high-resolution BTF into low-resolution images as produced by the BTF acquisition setup, similar to the simple down-sampling matrix used in the classical single-image super-resolution case referenced above. In general, the resulting linear system is under-determined and involves huge sparse matrices with unfavorable sparsity structure. However, by assuming priors such as the ones considered in the present thesis, such as that BTFs are low-rank, or that they are representable in terms of a global BTF model, or that they have roughly similar singular values, and by applying numerical tricks to cut down the size of the problem, it might actually become tractable. Preliminary experiments I conducted on synthetic data support this hypothesis; however, time did not permit to make the real-world case work. If it can in fact be made to work, it might also be possible to combine super-resolution with sparse and multiplexed acquisition into one single, elegant problem, providing a single, practical solution to most shortcomings of common types of

BTF acquisition setups at once. I consider this direction of research particularly worthwhile, as it relies mostly on the underlying physics and not on AI "hallucinations" as with our single-image super-resolution algorithm.

Finally, as of this writing, it seems that "deep learning" has found its way into most branches of science, with computer graphics and vision being no exception, of course, and the ongoing "revolution" does not appear to be ending anytime soon, given that it is ubiquitous and regularly produces results that would have been unthinkable not too many years ago. With the advent of neural network models for BTFs [83, 118, 119], it is likely that we will see another qualitative jump in the context of BTFs, irrespective of the particular problem. Regardless of what is actually possible with this first iteration of neural BTF models, I am certain that at some point there will be an iteration that can be used for the same or similar purposes as the models introduced in the present thesis. Given how well common networks lend themselves to image processing, it is conceivable that in the future, it may be possible to combine all the methods presented above into a single algorithm which produces a high-quality, high-resolution BTF from low-resolution, (learned) illumination-multiplexed input, or even from short snippets of low-dynamic-range smartphone video under uncontrolled illumination. By incorporating knowledge of the underlying physics as some kind of prior into the algorithm, inconsistencies from neural "hallucinations" could possibly be avoided. However, as Rainer et al. note themselves in their publication on neural BTF models [118], training neural networks robustly requires a lot of training data. The above methods should be able to play an important role in acquiring the necessary amounts of training data in a reasonable amount of time and fit them into a reasonable amount of memory. Moreover, at the moment they represent the benchmark when it comes to "computational" BTF acquisition.

In summary, the present thesis contains a variety of pioneering methods which make modelling and capturing digital material appearance in the form of bidirectional texture functions a lot more practical by significantly reducing the amount of time and data required to obtain a digital representation of a given material. Even though they only represent a first step in certain directions and will, at some point, likely be superseded by methods based on modern "deep learning" techniques, they are usable right now and can serve as a foundation for acquiring the required vast amounts of training data for more advanced models and techniques in the future.

Bibliography

[1] D. den Brok et al., *Patch-Based Sparse Reconstruction of Material BTFs*, Journal of WSCG **22** (2014) 83, ISSN: ISSN 1213-6972 (cit. on pp. 14, 18, 19, 34).

[2] D. den Brok et al., "Multiplexed Acquisition of Bidirectional Texture Functions for Materials," *Measuring, Modeling, and Reproducing Material Appearance II (SPIE 9398)*, vol. 9398, 2015 (cit. on pp. 14, 15, 22, 24).

[3] D. den Brok, H. C. Steinhausen, and R. Klein, "Fast multiplexed acquisition of high-dynamic-range material appearance," *Vision, Modeling & Visualization*, ed. by D. Bommes, T. Ritschel, and T. Schultz, Eurographics Association, 2015 151, ISBN: 978-3-905674-95-8 (cit. on pp. 15, 22, 24).

[4] D. den Brok, M. Weinmann, and R. Klein, *Rapid material capture through sparse and multiplexed measurements*, Computers & Graphics **73** (2018) 26, ISSN: 0097-8493, URL: `http://www.sciencedirect.com/science/article/pii/S0097849318300360` (cit. on pp. 15, 26).

[5] D. den Brok et al., "Per-Image Super-Resolution for Material BTFs," *2020 IEEE International Conference on Computational Photography (ICCP)*, 2020 1 (cit. on p. 15).

[6] H. C. Steinhausen et al., *Acquiring Bidirectional Texture Functions for Large-Scale Material Samples*, Journal of WSCG **22** (2014) 73 (cit. on pp. 15, 35).

[7] H. C. Steinhausen et al., "Extrapolation of bidirectional texture functions using texture synthesis guided by photometric normals," *Measuring, Modeling, and Reproducing Material Appearance 2015*, ed. by M. V. O. Segovia, P. Urban, and F. H. Imai, vol. 9398, International Society for Optics and Photonics, SPIE, 2015 81, URL: `https://doi.org/10.1117/12.2075717` (cit. on p. 15).

[8] D. den Brok, M. Weinmann, and R. Klein, "Linear Models for Material BTFs," *Eurographics Workshop on Material Appearance Modeling*, ISSN: 2309-5059, The Eurographics Association, 2015 15, ISBN: 978-3-905674-83-5 (cit. on p. 15).

[9] H. C. Steinhausen et al.,
"Extrapolating Large-Scale Material BTFs under Cross-Device Constraints,"
Vision, Modeling & Visualization, ed. by D. Bommes, T. Ritschel, and T. Schultz,
The Eurographics Association, 2015, ISBN: 978-3-905674-95-8 (cit. on pp. 15, 35).

[10] M. Weinmann et al., "Appearance Capture and Modeling,"
SIGGRAPH Asia 2015 Courses, SA '15,
Association for Computing Machinery, 2015, ISBN: 9781450339247,
URL: https://doi.org/10.1145/2818143.2835226 (cit. on p. 15).

[11] D. den Brok, M. Weinmann, and R. Klein,
"Towards Sparse and Multiplexed Acquisition of Material BTFs,"
Workshop on Material Appearance Modeling, ed. by R. Klein and H. Rushmeier,
The Eurographics Association, 2017, ISBN: 978-3-03868-035-2 (cit. on p. 15).

[12] H. C. Steinhausen et al., "Data-Driven Enhancement of SVBRDF Reflectance Data,"
Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Volume 1: GRAPP, 2018
(cit. on pp. 15, 33, 35).

[13] C. Callenberg et al., "EMCCD-SPAD Camera data fusion for high spatial resolution time-of-flight imaging.,"
Imaging and Applied Optics 2019 (COSI, IS, MATH, pcAOP),
Optical Society of America, 2019 CTh2A.3, URL:
http://www.osapublishing.org/abstract.cfm?URI=COSI-2019-CTh2A.3
(cit. on p. 16).

[14] C. Callenberg et al.,
Super-resolution time-resolved imaging using computational sensor fusion,
Scientific Reports **11** (2021) 1689, ISSN: 2045-2322,
URL: https://doi.org/10.1038/s41598-021-81159-x (cit. on p. 16).

[15] K. J. Dana et al., *Reflectance and Texture of Real-world Surfaces*,
ACM Trans. Graph. **18** (1999) 1, Place: New York, NY, USA Publisher: ACM,
ISSN: 0730-0301, URL: http://doi.acm.org/10.1145/300776.300778.

[16] Y. Dong et al., *Manifold Bootstrapping for SVBRDF Capture*,
ACM Trans. Graph. **29** (2010) 98:1, Place: New York, NY, USA Publisher: ACM,
ISSN: 0730-0301, URL: http://doi.acm.org/10.1145/1778765.1778835.

[17] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory **52** (2006) 1289.

[18] C. Eckart and G. Young,
The approximation of one matrix by another of lower rank, English,
Psychometrika **1** (1936) 211, Publisher: Springer-Verlag, ISSN: 0033-3123,
URL: http://dx.doi.org/10.1007/BF02288367 (cit. on p. 9).

[19] J. Filip et al.,
A Psychophysically Validated Metric for Bidirectional Texture Data Reduction,
ACM Trans. Graph. **27** (2008) 138:1, Place: New York, NY, USA Publisher: ACM,
ISSN: 0730-0301, URL: http://doi.acm.org/10.1145/1409060.1409091.

[20] P. Hall, D. Marshall, and R. Martin, *Merging and Splitting Eigenspace Models*,
IEEE Trans. Pattern Anal. Mach. Intell. **22** (2000) 1042, Place: Washington, DC,
USA Publisher: IEEE Computer Society, ISSN: 0162-8828,
URL: http://dx.doi.org/10.1109/34.877525.

[21] J. Y. Han and K. Perlin,
Measuring Bidirectional Texture Reflectance with a Kaleidoscope,
ACM Trans. Graph. **22** (2003) 741, Place: New York, NY, USA Publisher: ACM,
ISSN: 0730-0301, URL: http://doi.acm.org/10.1145/882262.882341
(cit. on p. 11).

[22] K. Marwah et al., *Compressive Light Field Photography Using Overcomplete Dictionaries and Optimized Projections*,
ACM Trans. Graph. **32** (2013) 46:1, Place: New York, NY, USA Publisher: ACM,
ISSN: 0730-0301, URL: http://doi.acm.org/10.1145/2461912.2461914.

[23] W. Matusik et al., *A Data-driven Reflectance Model*,
ACM Trans. Graph. **22** (2003) 759, Place: New York, NY, USA Publisher: ACM,
ISSN: 0730-0301, URL: http://doi.acm.org/10.1145/882262.882343
(cit. on pp. 8, 18, 25).

[24] W. Matusik et al., "Efficient Isotropic BRDF Measurement,"
Proceedings of the 14th Eurographics Workshop on Rendering, EGRW '03,
event-place: Leuven, Belgium, Eurographics Association, 2003 241,
ISBN: 3-905673-03-7,
URL: http://dl.acm.org/citation.cfm?id=882404.882439
(cit. on pp. 8, 11).

[25] G. Müller, J. Meseth, and R. Klein,
"Compression and Real-Time Rendering of Measured BTFs Using Local PCA,"
Vision, Modeling and Visualisation 2003,
Akademische Verlagsgesellschaft Aka GmbH, Berlin, 2003 271,
ISBN: 3-89838-048-3.

[26] G. Müller, G. H. Bendels, and R. Klein, "Rapid Synchronous Acquisition of Geometry and BTF for Cultural Heritage Artefacts," *The 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST)*,
Backup Publisher: Eurographics Association, Eurographics Association, 2005 13
(cit. on p. 12).

[27] P. Peers et al., *Compressive Light Transport Sensing*,
ACM Trans. Graph. **28** (2009) 3:1, Place: New York, NY, USA Publisher: ACM,
ISSN: 0730-0301, URL: http://doi.acm.org/10.1145/1477926.1477929.

[28] S. Roweis, "EM Algorithms for PCA and SPCA," *Proceedings of the 1997 Conference on Advances in Neural Information Processing Systems 10*, NIPS '97, event-place: Denver, Colorado, USA, MIT Press, 1998 626, ISBN: 0-262-10076-2, URL: http://dl.acm.org/citation.cfm?id=302528.302762.

[29] C. Schwartz et al.,
Design and Implementation of Practical Bidirectional Texture Function Measurement Devices focusing on the Developments at the University of Bonn, Sensors **14** (2014), ISSN: 1424-8220,
URL: http://www.mdpi.com/1424-8220/14/5/7753 (cit. on pp. 13, 21).

[30] C. Schwartz et al., "DOME II: A Parallelized BTF Acquisition System,"
Eurographics Workshop on Material Appearance Modeling: Issues and Acquisition, event-place: Zaragoza, Spain, Eurographics Association, 2013 25,
ISBN: 978-3-905674-48-4,
URL: http://diglib.eg.org/EG/DL/WS/MAM/MAM2013/025-031.pdf
(cit. on pp. 12, 23).

[31] M. Varma and A. Zisserman,
A Statistical Approach to Material Classification Using Image Patch Exemplars, IEEE Trans. Pattern Anal. Mach. Intell. **31** (2009) 2032, Place: Washington, DC, USA Publisher: IEEE Computer Society, ISSN: 0162-8828,
URL: http://dx.doi.org/10.1109/TPAMI.2008.182.

[32] T.-T. Wong et al., "Image-based Rendering with Controllable Illumination,"
Proceedings of the Eurographics Workshop on Rendering Techniques '97, Springer-Verlag, 1997 13, ISBN: 3-211-83001-4,
URL: http://dl.acm.org/citation.cfm?id=647651.731971.

[33] M. Harwit and N. Sloane, *Hadamard transform optics*, Academic Press, 1979,
ISBN: 978-0-12-330050-8,
URL: http://books.google.de/books?id=UAy1AAAAIAAJ (cit. on pp. 22, 23).

[34] M. L. Koudelka et al.,
"Acquisition, compression, and synthesis of bidirectional texture functions,"
ICCV Workshop on Texture Analysis and Synthesis, 2003.

[35] E. Miandji, J. Kronander, and J. Unger,
"Compressive Image Reconstruction in Reduced Union of Subspaces,"
Eurographics 2015, 2015.

[36] K. Mitra, O. Cossairt, and A. Veeraraghavan, "Can we beat Hadamard multiplexing? Data driven design and analysis for computational imaging systems,"
Computational Photography (ICCP), 2014 IEEE International Conference on, 2014 1.

[37] N. Ratner and Y. Schechner,
"Illumination Multiplexing within Fundamental Limits,"
Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on,
ISSN: 1063-6919, 2007 1.

[38] Y. Y. Schechner, S. K. Nayar, and P. N. Belhumeur,
Multiplexing for Optimal Lighting, IEEE Trans. Pattern Anal. Mach. Intell. **29**
(2007) 1339, Place: Washington, DC, USA Publisher: IEEE Computer Society,
ISSN: 0162-8828, URL: http://dx.doi.org/10.1109/TPAMI.2007.1151.

[39] A. Wenger et al., *Performance Relighting and Reflectance Transformation with
Time-multiplexed Illumination*,
ACM Trans. Graph. **24** (2005) 756, Place: New York, NY, USA Publisher: ACM,
ISSN: 0730-0301, URL: http://doi.acm.org/10.1145/1073204.1073258.

[40] M. Aittala, T. Weyrich, and J. Lehtinen,
Practical SVBRDF Capture in the Frequency Domain,
ACM Trans. Graph. **32** (2013) 110:1, Place: New York, NY, USA Publisher: ACM,
ISSN: 0730-0301, URL: http://doi.acm.org/10.1145/2461912.2461978.

[41] M. Aittala, T. Aila, and J. Lehtinen,
Reflectance Modeling by Neural Texture Synthesis,
ACM Trans. Graph. **35** (2016) 65:1, Place: New York, NY, USA Publisher: ACM,
ISSN: 0730-0301, URL: http://doi.acm.org/10.1145/2897824.2925917.

[42] M. Aittala, T. Weyrich, and J. Lehtinen,
Two-shot SVBRDF Capture for Stationary Materials,
ACM Trans. Graph. **34** (2015) 110:1, Place: New York, NY, USA Publisher: ACM,
ISSN: 0730-0301, URL: http://doi.acm.org/10.1145/2766967 (cit. on p. 11).

[43] M. M. Bagher, J. Snyder, and D. Nowrouzezahrai,
A Non-Parametric Factor Microfacet Model for Isotropic BRDFs,
ACM Trans. Graph. **35** (2016) 159:1, Place: New York, NY, USA Publisher: ACM,
ISSN: 0730-0301, URL: http://doi.acm.org/10.1145/2907941.

[44] A. Davis, M. Levoy, and F. Durand, *Unstructured Light Fields*,
Comput. Graph. Forum **31** (2012) 305, Place: Chichester, UK Publisher: The
Eurographs Association & John Wiley & Sons, Ltd., ISSN: 0167-7055,
URL: http://dx.doi.org/10.1111/j.1467-8659.2012.03009.x.

[45] J. Filip and M. Haindl,
Bidirectional Texture Function Modeling: A State of the Art Survey,
IEEE Trans. Pattern Anal. Mach. Intell. **31** (2009) 1921.

[46] M. Haindl and J. Filip, *Visual Texture: Accurate Material Appearance Measurement,
Representation and Modeling*,
Advances in Computer Vision and Pattern Recognition, Springer London, 2013,
ISBN: 9781447149026,

URL: https://books.google.de/books?id=%5C_R4%5C_AAAAQBAJ
(cit. on pp. 3, 4).

[47] F. L. Hitchcock, *The Expression of a Tensor or a Polyadic as a Sum of Products*, Journal of Mathematics and Physics **6** (1927) 164, ISSN: 1467-9590, URL: http://dx.doi.org/10.1002/sapm192761164.

[48] J. Yu et al., "Sparse Sampling for Image-Based SVBRDF Acquisition," *Workshop on Material Appearance Modeling*, ed. by R. Klein and H. Rushmeier, ISSN: 2309-5059, The Eurographics Association, 2016, ISBN: 978-3-03868-007-9.

[49] C.-K. Liang et al., *Programmable Aperture Photography: Multiplexed Light Field Acquisition*, ACM Trans. Graph. **27** (2008) 55:1, Place: New York, NY, USA Publisher: ACM, ISSN: 0730-0301, URL: http://doi.acm.org/10.1145/1360612.1360654.

[50] N. Menzel and M. Guthe, "High Dynamic Range Preserving Compression of Light Fields and Reflectance Fields," *Proceedings of the 5th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*, AFRIGRAPH '07, Association for Computing Machinery, 2007 71, ISBN: 9781595939067, URL: https://doi.org/10.1145/1294685.1294697 (cit. on p. 25).

[51] I. Miyagawa and Y. Taniguchi, *Dense Light Transport for Relighting Computation Using Orthogonal Illumination Based on Walsh-Hadamard Matrix*, IEICE Transactions on Information and Systems **E99-D** (2016) 1038, Publisher: IEICE, ISSN: 1745-1361.

[52] G. Müller et al., "Acquisition, Synthesis and Rendering of Bidirectional Texture Functions," *Eurographics 2004, State of the Art Reports*, 2004 69.

[53] F. E. Nicodemus et al., *Geometrical considerations and nomenclature for reflectance*, National Bureau of Standards Monograph #160, U.S. Department of Commerce, 1977.

[54] J. B. Nielsen, H. W. Jensen, and R. Ramamoorthi, *On Optimal, Minimal BRDF Sampling for Reflectance Acquisition*, ACM Trans. Graph. **34** (2015) 186:1 (cit. on p. 26).

[55] J. I. Park et al., "Multispectral Imaging Using Multiplexed Illumination," *International Conference on Computer Vision*, ISSN: 1550-5499, 2007 1.

[56] R. Ruiters, C. Schwartz, and R. Klein, *Data Driven Surface Reflectance from Sparse and Irregular Samples*, Computer Graphics Forum (Proc. of Eurographics) **31** (2012) 315.

[57] S. M. Rusinkiewicz,
"A New Change of Variables for Efficient BRDF Representation,"
Rendering Techniques '98: Proceedings of the Eurographics Workshop in Vienna, Austria, June 29—July 1, 1998, ed. by G. Drettakis and N. Max,
Springer Vienna, 1998 11, ISBN: 978-3-7091-6453-2,
URL: http://dx.doi.org/10.1007/978-3-7091-6453-2_2 (cit. on p. 8).

[58] N. Thanikachalam et al., *Scene Relighting with Smartphones*,
IEEE Transactions on Computational Imaging (2016),
URL: http://go.epfl.ch/relightables.

[59] R. Vavra and J. Filip,
Minimal Sampling for Effective Acquisition of Anisotropic BRDFs,
Computer Graphics Forum (PACIFIC GRAPHICS 2016) (2016) 299, Publisher: The Eurographics Association and John Wiley & Sons Ltd.,
URL: http://staff.utia.cas.cz/filip/projects/16PG.

[60] M. Weinmann and R. Klein,
"Advances in Geometry and Reflectance Acquisition (Course Notes),"
SIGGRAPH Asia 2015 Courses, SA '15, event-place: Kobe, Japan, ACM, 2015 1:1,
ISBN: 978-1-4503-3924-7,
URL: http://doi.acm.org/10.1145/2818143.2818165 (cit. on pp. 3, 11).

[61] Z. Xu et al.,
Minimal BRDF Sampling for Two-Shot Near-Field Reflectance Acquisition, ACM Trans. Graph. **35** (2016), Publisher: Association for Computing Machinery, Inc.,
ISSN: 0730-0301.

[62] Z. Zhou et al., *Sparse-as-possible SVBRDF Acquisition*,
ACM Trans. Graph. **35** (2016) 189:1, Place: New York, NY, USA Publisher: ACM,
ISSN: 0730-0301, URL: http://doi.acm.org/10.1145/2980179.2980247.

[63] S. Anwar, S. Khan, and N. Barnes, *A Deep Journey into Super-resolution: A survey*,
arXiv preprint arXiv:1904.07523 (2019).

[64] F. Cao et al., *Image Super-Resolution via Adaptive $\ell _ ${$p$}$(0< p< 1)$ Regularization and Sparse Representation*, IEEE transactions on neural networks and learning systems **27** (2016) 1550, Publisher: IEEE.

[65] D. Dai, R. Timofte, and L. Van Gool,
"Jointly Optimized Regressors for Image Super-resolution,"
Computer Graphics Forum, vol. 34, Issue: 2, Wiley Online Library, 2015 95.

[66] S. Dai et al., *Softcuts: a soft edge smoothness prior for color image super-resolution*,
IEEE Transactions on Image Processing **18** (2009) 969, Publisher: IEEE.

[67] K. J. Dana et al., "Reflectance and Texture of Real World Surfaces,"
IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1997 151.

[68] W. Dong et al., *Bidirectional texture function image super-resolution using singular value decomposition*,
Applied optics **56** (2017) 2745, Publisher: Optical Society of America (cit. on p. 29).

[69] C. Dong et al., *Image Super-Resolution Using Deep Convolutional Networks*,
IEEE Trans. Pattern Anal. Mach. Intell. **38** (2016) 295, Place: Washington, DC, USA
Publisher: IEEE Computer Society, ISSN: 0162-8828,
URL: http://dx.doi.org/10.1109/TPAMI.2015.2439281.

[70] C. Dong et al.,
"Learning a Deep Convolutional Network for Image Super-Resolution,"
Computer Vision – ECCV 2014, ed. by D. Fleet et al.,
Springer International Publishing, 2014 184, ISBN: 978-3-319-10593-2.

[71] C. E. Duchon, *Lanczos Filtering in One and Two Dimensions.*,
Journal of Applied Meteorology **18** (1979) 1016.

[72] W. T. Freeman, T. R. Jones, and E. C. Pasztor, *Example-based super-resolution*,
IEEE Computer graphics and Applications (2002) 56, Publisher: IEEE.

[73] Y. Gan et al., "Perception driven texture generation,"
2017 IEEE International Conference on Multimedia and Expo (ICME),
ISSN: 1945-788X, 2017 889.

[74] S. Ioffe and C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*,
CoRR **abs/1502.03167** (2015), _eprint: 1502.03167,
URL: http://arxiv.org/abs/1502.03167.

[75] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*,
abs/1412.6980 (2014), _eprint: 1412.6980,
URL: http://arxiv.org/abs/1402.6980.

[76] W.-S. Lai et al.,
"Deep laplacian pyramid networks for fast and accurate super-resolution,"
Proceedings of the IEEE conference on computer vision and pattern recognition,
2017 624.

[77] W.-S. Lai et al.,
Fast and accurate image super-resolution with deep laplacian pyramid networks,
IEEE transactions on pattern analysis and machine intelligence (2018), Publisher: IEEE.

[78] C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network,"
2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),
ISSN: 1063-6919, IEEE Computer Society, 2017 105,
URL: https://doi.ieeecomputersociety.org/10.1109/CVPR.2017.19.

[79] J. Liu et al.,
Retrieval compensated group structured sparsity for image super-resolution,
IEEE Transactions on Multimedia **19** (2016) 302, Publisher: IEEE.

[80] A. Marquina and S. J. Osher,
Image super-resolution by TV-regularization and Bregman iteration,
Journal of Scientific Computing **37** (2008) 367, Publisher: Springer.

[81] M. Mirza and S. Osindero, *Conditional Generative Adversarial Nets*, 2014,
URL: http://arxiv.org/abs/1411.1784.

[82] S.-J. Park et al.,
"SRFeat: Single Image Super-Resolution with Feature Discrimination,"
Computer Vision – ECCV 2018, ed. by V. Ferrari et al.,
Springer International Publishing, 2018 455, ISBN: 978-3-030-01270-0.

[83] G. Rainer et al., *Neural BTF Compression and Interpolation*,
Computer Graphics Forum (Proceedings of Eurographics) **38** (2019)
(cit. on pp. 34, 37).

[84] M. A. Robertson, S. Borman, and R. L. Stevenson,
"Dynamic range improvement through multiple exposures,"
Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348),
vol. 3, ISSN: null, 1999 159 (cit. on p. 12).

[85] M. S. Sajjadi, B. Scholkopf, and M. Hirsch,
"Enhancenet: Single image super-resolution through automated texture synthesis,"
Proceedings of the IEEE International Conference on Computer Vision, 2017 4491.

[86] W. Shi et al., "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,"
Proceedings of the IEEE conference on computer vision and pattern recognition,
2016 1874.

[87] T. A. Stephenson and T. Chen,
Adaptive Markov random fields for example-based super-resolution of faces,
EURASIP Journal on Advances in Signal Processing **2006** (2006) 031062, Publisher: Springer.

[88] J. Sun, Z. Xu, and H.-Y. Shum, "Image super-resolution using gradient profile prior,"
2008 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2008.

[89] R. Timofte, V. De Smet, and L. Van Gool,
"A+: Adjusted anchored neighborhood regression for fast super-resolution,"
Asian conference on computer vision, Springer, 2014 111.

[90] R. Timofte, V. De Smet, and L. Van Gool,
"Anchored neighborhood regression for fast example-based super-resolution,"
Proceedings of the IEEE international conference on computer vision, 2013 1920.

[91] T. Tong et al., "Image Super-Resolution Using Dense Skip Connections," *2017 IEEE International Conference on Computer Vision (ICCV)*, ISSN: 2380-7504, 2017 4809.

[92] Z. Wang et al., "Deep networks for image super-resolution with sparse prior," *Proceedings of the IEEE international conference on computer vision*, 2015 370.

[93] X. Wang et al., *ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks*, CoRR **abs/1809.00219** (2018), _eprint: 1809.00219, URL: http://arxiv.org/abs/1809.00219.

[94] B. Wu et al., *SRPGAN: Perceptual Generative Adversarial Network for Single Image Super Resolution*, CoRR **abs/1712.05927** (2017), URL: http://arxiv.org/abs/1712.05927.

[95] Q. Yan et al., *Single image superresolution based on gradient profile sharpness*, IEEE Transactions on Image Processing **24** (2015) 3187, Publisher: IEEE.

[96] W. Yang et al., *Deep Learning for Single Image Super-Resolution: A Brief Review*, ArXiv **abs/1808.03344** (2018).

[97] W. Yifan et al., "A Fully Progressive Approach to Single-Image Super-Resolution," *CVPR Workshops*, 2018.

[98] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," *International conference on curves and surfaces*, Springer, 2010 711.

[99] X. Zhang et al., *BTF data Generation based on Deep Learning*, Procedia Computer Science **147** (2019) 233, ISSN: 1877-0509, URL: http://www.sciencedirect.com/science/article/pii/S1877050919302637.

[100] Y. Zhang et al., "Image super-resolution based on dictionary learning and anchored neighborhood regression with mutual incoherence," *2015 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2015 591.

[101] Y. Zhang et al., "Residual Dense Network for Image Super-Resolution," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, ISSN: 1063-6919, 2018 2472.

[102] X. Zhang et al., "Zoom to Learn, Learn to Zoom," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[103] K. B. Meena and V. Tyagi, *Distinguishing computer-generated images from photographic images using two-stream convolutional neural network*, Applied Soft Computing **100** (2021) 107025, ISSN: 1568-4946, URL: https://www.sciencedirect.com/science/article/pii/S1568494620309649 (cit. on p. 1).

[104] M. Weinmann, J. Gall, and R. Klein,
"Material Classification Based on Training Data Synthesized Using a BTF Database,"
Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part III, Springer International Publishing, 2014 156 (cit. on p. 1).

[105] M. Rodrigo, W. Michael, and H. Matthias B.,
Digital Transmission of Subjective Material Appearance,
Journal of WSCG **25** (2017) 57, ISSN: 1213-6972,
URL: http://wscg.zcu.cz/WSCG2017/!_2017_Journal_WSCG-No-2.pdf
(cit. on p. 1).

[106] R. W. Fleming, *Material Perception*,
Annual Review of Vision Science **3** (2017) 365, PMID: 28697677,
eprint: https://doi.org/10.1146/annurev-vision-102016-061429,
URL: https://doi.org/10.1146/annurev-vision-102016-061429
(cit. on p. 3).

[107] J. T. Kajiya, "The Rendering Equation," *Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '86,
Association for Computing Machinery, 1986 143, ISBN: 0897911962,
URL: https://doi.org/10.1145/15922.15902 (cit. on p. 4).

[108] G. J. Ward, *Measuring and Modeling Anisotropic Reflection*,
SIGGRAPH Comput. Graph. **26** (1992) 265, ISSN: 0097-8930,
URL: https://doi.org/10.1145/142920.134078 (cit. on p. 8).

[109] A. Ngan, F. Durand, and W. Matusik, "Experimental Analysis of BRDF Models,"
Proceedings of the Eurographics Symposium on Rendering,
Eurographics Association, 2005 117 (cit. on p. 8).

[110] M. Kurt and D. Edwards, *A Survey of BRDF Models for Computer Graphics*,
SIGGRAPH Comput. Graph. **43** (2009), ISSN: 0097-8930,
URL: https://doi.org/10.1145/1629216.1629222 (cit. on p. 8).

[111] Y. Dong, *Deep appearance modeling: A survey*, Visual Informatics **3** (2019) 59,
ISSN: 2468-502X, URL: https://www.sciencedirect.com/science/article/pii/S2468502X1930035X
(cit. on pp. 8, 34).

[112] H. Wu, J. Dorsey, and H. Rushmeier,
A Sparse Parametric Mixture Model for BTF Compression, Editing and Rendering,
Computer Graphics Forum **30** (2011) 465,
eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8659.2011.01890.x,
URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8659.2011.01890.x (cit. on p. 9).

[113] M. Haindl, "Bidirectional Texture Function Modeling,"
Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging: Mathematical Imaging and Vision, ed. by K. Chen et al.,
Springer International Publishing, 2021 1, ISBN: 978-3-030-03009-4,
URL: https://doi.org/10.1007/978-3-030-03009-4_103-1 (cit. on p. 10).

[114] S. Merzbach et al., *Learned Fitting of Spatially Varying BRDFs*,
Computer Graphics Forum (2019), ISSN: 1467-8659 (cit. on p. 11).

[115] M. Schöberl et al., *Photometric limits for digital camera systems*,
Journal of Electronic Imaging **21** (2012) 020501,
URL: https://doi.org/10.1117/1.JEI.21.2.020501 (cit. on p. 13).

[116] V. Havran et al.,
Increasing the Spatial Resolution of BTF Measurement with Scheimpflug Imaging,
Computer Graphics Forum **38** (2019) 592,
eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.13594,
URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13594
(cit. on pp. 14, 33).

[117] K. Kang et al., *Learning Efficient Illumination Multiplexing for Joint Capture of Reflectance and Shape*, ACM Trans. Graph. **38** (2019), ISSN: 0730-0301,
URL: https://doi.org/10.1145/3355089.3356492 (cit. on p. 33).

[118] G. Rainer et al., "Unified neural encoding of BTFs," *Computer Graphics Forum*, vol. 39, 2, Wiley Online Library, 2020 167 (cit. on pp. 34, 37).

[119] A. Kuznetsov et al., "Rendering Neural Materials on Curved Surfaces,"
ACM SIGGRAPH 2022 Conference Proceedings, SIGGRAPH '22,
Association for Computing Machinery, 2022, ISBN: 9781450393379,
URL: https://doi.org/10.1145/3528233.3530721 (cit. on pp. 34, 37).

[120] T. Yao et al., "Single-Image Super-Resolution: A Survey,"
Communications, Signal Processing, and Systems, ed. by Q. Liang et al.,
Springer Singapore, 2020 119 (cit. on p. 34).

[121] S. M. A. Bashir et al.,
A comprehensive review of deep learning-based single image super-resolution,
PeerJ Computer Science **7** (2021) e621 (cit. on p. 34).

[122] R. Ruiters, C. Schwartz, and R. Klein,
Example-based Interpolation and Synthesis of Bidirectional Texture Functions,
Computer Graphics Forum **32** (2013) 361,
eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.12056,
URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12056
(cit. on p. 36).

Appendix

APPENDIX A

Patch-based sparse reconstruction of material BTFs

Patch-based sparse reconstruction of material BTFs

Dennis den Brok Heinz Christian Steinhausen Matthias Hullin Reinhard Klein

Universität Bonn
Institute of Computer Science II
Friedrich-Ebert-Allee 144
53113 Bonn, Germany
{ denbrok, steinhau, hullin, rk }@cs.uni-bonn.de

ABSTRACT

We propose a simple and efficient method to reconstruct materials' *bidirectional texture functions* (BTFs) from angularly sparse measurements. The key observation is that materials of similar types exhibit both similar surface structure and reflectance properties. We exploit this by manually clustering an existing database of fully measured material BTFs and fitting a linear model to each of the clusters. The models are computed not on per-texel data but on small spatial BTF patches we call *apparent BTFs*. Sparse reconstruction can then be performed by solving a linear least-squares problem without any regularization, using a per-cluster sampling strategy derived from the models. We demonstrate that our method is capable of faithfully reconstructing fully resolved BTFs from sparse measurements for a wide range of materials.

Keywords

bidirectional texture functions, sparse acquisition, material appearance

1 INTRODUCTION

In many applications, it is desirable or even imperative to reproduce a material's appearance faithfully and, possibly, in real-time. For a wide range of materials *bidirectional texture functions* (BTFs) – loosely speaking, an image-based variant of the better known *spatially varying bidirectional reflectance distribution functions* (SVBRDFs) – provide good reproduction quality, even at interactive frame rates. The acquisition of high-quality, high-resolution BTFs of real-world materials is, however, by many means expensive. In particular, measurement times of typically many hours per material make it very cumbersome to obtain large BTF databases, as pictures of the material to be measured have to be taken from many different viewing angles and under many different lighting conditions.

We propose a simple and efficient method for the sparse acquisition of material BTFs, assuming a sufficiently large and heterogenous database of fully measured materials is available:

We demonstrate that linear models describing material reflectance per texel are insufficient for this task because effects not local to texels frequently occur. We show that, instead, small BTF patches we call *apparent BTFs* (ABTFs) provide a suitable foundation for such models. In order to account for the high variance of material surfaces, we propose to fit models to patches clustered by semantic material class. From these models, sparse sampling strategies can be deduced that take advantage of the peculiarities of existing BTF acquisition devices. Reconstruction from such sparse measurements can then be achieved efficiently by solving a simple linear least-squares problem without regularization.

We demonstrate that our method is able to reconstruct fully resolved material BTFs of good quality from as little as 6% of the original samples. It can be used for substantially improving acquisition times or angular resolution, thus benefiting the most common BTF acquisition devices.

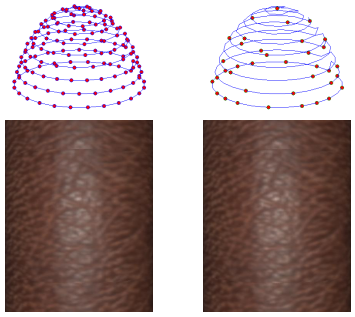


Figure 1: Left: Reference rendering along with the full sampling. Right: Sparse sampling used to produce the rendering of a sparse reconstruction shown below.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

2 BACKGROUND

2.1 Bidirectional texture functions

BTFs have been introduced by Dana *et al*. [1] as an image-based approach to spatially varying appearance. Like SVBRDFs, they are 6-dimensional functions of the form

$$\mathscr{B}(\mathbf{x}, \omega_i, \omega_o),$$

where $\omega_{i,o} \in \mathbf{R}^2$ are the incoming and outgoing light directions, respectively, and $\mathbf{x} \in \mathbf{R}^2$ is the position on a parameterized surface V. In the case of material BTFs, V is typically flat; it does not need to coincide with the material's actual surface geometry. It is generally assumed that light sources are directional and have the same spectrum. In particular, effects such as phosphorescene, fluorescence and subsurface scattering cannot be captured accurately.

The fundamental difference from SVBDRFs is that the function $\mathscr{B}(\mathbf{x}, -)$ need not be BRDF-valued: the corresponding per-texel reflectance function does not need to adhere to Helmholtz reciprocity and conservation of energy and is, therefore, capable of capturing non-local effects such as interreflections and self-shadowing. Moreover, because V does not necessarily coincide with the actual surface, the per-texel reflectance functions may also describe parallax effects. For these reasons, the term *apparent BRDF* (ABRDF) has been suggested by Wong *et al*. [18] for this kind of functions. Conversely, the values of the function $\mathscr{B}(-, \omega_i, \omega_o)$ are just 2D textures corresponding to specific pairs of incoming and outgoing light directions.

2.2 BTF acquisition

Several setups for the acquisition of BTFs have been proposed. We briefly review the most prominent paradigms, as our method benefits all of them to a greater or lesser extent. An in-depth overview can be found in [16].

2.2.1 Gonioreflectometer

In what is historically the first BTF acquisition setup, proposed by Dana *et al*. [1], the material sample is placed on a turntable, and a camera and a light source held by robot arms are moved across the hemisphere above the sample to capture images of the sample under different lighting and viewing conditions. The gonioreflectometer is very flexible in terms of possible samplings of the hemisphere, but measurement times are excessive – on the order of weeks for a moderate angular resolution – due to the little amounts of light sources and sensors and the movable parts' low speeds.

2.2.2 Kaleidoscope

Han *et al*. [7] introduced an intriguing parallel setup: The sample is placed underneath a tapered kaleidoscope, lit and captured from a projector and a camera placed at the other end, which allows for a number of lighting and viewing conditions to be measured in a single camera shot. By appropriately arranging the mirrors, the angular and spatial resolution can be adjusted; however, both are typically rather low, and increasing one leads to a decrease of the other, so there is always a tradeoff to be made.

2.2.3 Camera domes

Camera domes as proposed, for instance, by Müller *et al*. [11] and Schwartz *et al*. [15] ideally provide a highly parallel means to acquire BTFs: A number of cameras is spread across the hemisphere above the sample holder. Their flashes or separate LEDs are used as light sources. Parallelism may be traded for fewer cameras and lower cost by placing the sample on a turntable in order to achieve a similarly dense sampling of the hemisphere. Due to the number of cameras, data transfer times become a new bottleneck.

In all of the above setups, it is usually necessary to capture the same scene several times with different shutter times in order to obtain HDR data.

2.3 Related work

To the best of our knowledge, no method for sparse reconstruction of entire BTFs has been proposed so far. There exists, however, a number of methods for lower-dimensional reflectance models:

In [10], Matusik *et al*. perform *singular value decomposition* (SVD) on a database of 100 measured BRDFs of a wide range of isotropic materials to obtain a linear model.

In [9], the same authors introduce two methods for sparse reconstruction of isotropic BRDFs: The first method is based on a wavelet analysis of their BRDF database. A set of basis wavelets termed *common wavelet basis* is determined and used to reconstruct previously unseen BRDFs with approximately 1.5 million samples from approximately 70000 measurements. The second method uses the entire BRDF database itself as a linear model for reconstruction of fully measured BRDFs from as little as 800 out of the original approximately 1.5 million samples, at the cost of slightly increased reconstruction errors and the required availability of the BRDF database. Samples are chosen using a simple optimization algorithm such that the linear system to be solved for reconstruction is well-conditioned. They do not investigate how well their methods generalize to more complex reflectance such as anisotropic BRDFs or ABRDFs.

In [2], Dong *et al*. reconstruct a material's SVBRDF from a sparse measurement using a manifold constructed from analytical BRDFs fit to fully measured BRDFs of manually selected representative points on

the material's surface. The algorithm is unlikely to scale to BTFs because of the typically much higher intrinsic dimensionality of the ABRDF manifold (*cf.* section 4.1.1). A generalization to previously unseen materials is not obvious, albeit conceivable.

Peers *et al.* [13] introduce *compressed sensing* [3] to the acquisition of *reflectance fields*, assuming both 2D *outgoing* (here: fixed viewing direction) and *incident light fields*. Their algorithm uses a hierarchical, multi-resolution Haar wavelet basis, taking spatial coherence into account. It is not clear how to extend this approach to a multi-view setup. Common BTF acquisition setups only have a very limited number of light sources, where the advantage of compressed sensing might be negligible. Due to these light sources' brightness, we expect shot noise to become a problem.

Conversely, in [8], Marwah *et al.* use sparsity-based methods related to compressed sensing in order to sparsely acquire 4D light fields with an angular resolution of 5×5. They compute a dictionary of what they call *light field atoms* – 11×11 spatial light field patches which allow for a sparse representation of natural light fields. Such a dictionary does not exist in the case of ABRDFs or ABTFs; as demonstrated in section 4.1.1, their dimensionality is likely too high.

Filip *et al.* [5] propose a vector quantization of BTFs for the purpose of compression, guided by a psychophysically validated metric. They conclude that as little as 10 – 35 % of the original textures are sufficient to maintain the same visual appearance in renderings. It would be interesting to investigate whether there is a common quantization for *all* materials, and if so, whether it could be used for sparse acquisition. A large user study would be needed in order to adapt the metric to a bigger BTF database.

3 LINEAR MODELS FOR MATERIAL BTFS

During measurement, a finite discretization of the measured material's BTF $\mathscr{B}$ is obtained. After rectification of the acquired images, the discrete BTF has a natural representation as a matrix $\mathbf{B} \in \mathbf{R}^{n \times m}$ with the columns representing the m discrete ABRDFs, each entry corresponding to some pair (ω_i, ω_o) of incoming and outgoing angle, and the rows representing the n rectified textures (*cf.* figure 2).

3.1 Linear models & reconstruction

The goal is now to recover $\mathbf{B}$ from a sparse measurement

$$\widetilde{\mathbf{B}} = \mathbf{M}\mathbf{B}$$

of n_s samples, where $\mathbf{M} \in \mathbf{R}^{n_s \times n}$ is a *measurement matrix*, typically binary, which determines the sparse sampling.

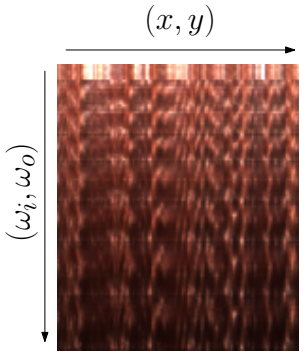


Figure 2: Representation of a discretized BTF as a matrix.

Arguably one of the most simple methods to attack this problem is, given fully resolved training data $\mathbf{D}$, to fit a linear model $\mathbf{D} \approx \mathbf{UC}$. An optimal fitting method in terms of L^2 error is to compute a truncated SVD

$$\mathbf{D} \approx \mathbf{U}\Sigma\mathbf{V}^t$$

as established by the Eckart-Young theorem [4]. The hope is that the model both

- generalizes to previously unknown data; *i.e.*

$$\min_{\mathbf{C_B}} \|\mathbf{UC_B} - \mathbf{B}\| < \varepsilon \tag{1}$$

- is expressive enough that a sparse sampling is sufficient to find reasonable coefficients; *i.e.*

$$\min_{\mathbf{C_B}} \|\mathbf{M}(\mathbf{UC_B} - \mathbf{B})\| < \delta \implies \|\mathbf{UC_B} - \mathbf{B}\| < \varepsilon \tag{2}$$

Provided $\widetilde{\mathbf{B}}$ has at least as many rows as columns, an approximation of $\mathbf{B}$ may then be obtained via

$$\mathbf{B} \approx \mathbf{U}(\mathbf{MU})^{\dagger}\widetilde{\mathbf{B}}, \tag{3}$$

where $(\mathbf{MU})^{\dagger}$ denotes the Moore-Penrose pseudo-inverse of $\mathbf{MU}$.

It is well-known that the fitting of linear models through minimization of L^2 error is sensitive to outliers. In order to decrease the influence of specular highlights, we reduce the data's dynamic range by converting the measured HDR RGB data to YUV color space, dividing the U and V values by the corresponding Y value and applying log to the Y values.

Despite its simplicity, this approach has been demonstrated in [9] to be quite effective in the special case of isotropic BRDFs. It seems thus worthwhile to investigate whether this generalizes to ABRDFs.

3.2 Linear models for ABRDFs

Linear models for ABRDFs are already being used for compression and rendering of BTFs, often under the

moniker *full matrix factorization* (FMF). In that case, models are fit to a certain material's ABRDFs only; *i.e.* to $\mathbf{B}$ instead of a whole database $\mathbf{D}$. The columns of $\mathbf{U}$ and $\mathbf{V}$ are commonly refered to as *eigen-ABRDFs* and *eigentextures*, respectively [12], in reference to their semantic meaning.

It is reasonable to assume that for BTFs the ABRDFs of which are close to being true BRDFs, a linear model may perform similarly well as in [9] [10] with respect to equations 1 and 2. However, as soon as surface structure becomes significant, reconstructions from sparse measurements might easily miss effects such as self-shadowing, interreflections, occlusion and parallax. We shall demonstrate in section 4.2 that this is indeed the case.

3.3 Linear models for ABTFs

In order to overcome these problems, we take spatial information into account: instead of considering only ABRDFs, we consider entire collections of neighboring ABRDFs, which we call *apparent BTFs* (ABTFs), as similar to ABRDFs they capture effects not local to the specific patch such as interreflections or shadows cast from neighboring patches.

The matrix $\mathbf{B}$ then takes on a different form, with its columns representing discrete ABTFs, for instance as vectors of stacked discrete ABRDFs belonging to the same neighborhood. The corresponding measurement matrix becomes $\mathbf{1}_{p^2} \otimes \mathbf{M}$, where p denotes the spatial patch size and $\otimes$ the Kronecker product.

Note that an alternative to patches exists in the form of appropriate filter banks, as *e.g.* demonstrated by Peers *et al.* [13] A case has been made in favor of the simpler spatial patches by Varma *et al.* [17], albeit in the case of material classification: the authors demonstrate that classification using spatial patches, which can be as small as 3×3 texels, is superior to that using filter banks with equivalent support.

The intrinsic dimensionality of the ABTF database is likely higher than that of the ABRDF database; in the worst case by a factor equal to the patch size. To mitigate this to some extent, we propose to cluster the database such that each cluster contains only materials with similar surface structure, and determine the linear models $\mathbf{D}_{\text{cluster}} \approx \mathbf{U}\Sigma\mathbf{V}^t$ per cluster. The columns of $\mathbf{U}$ shall be called *eigen-ABTFs*.

3.4 Sampling strategies

Once a model satisfying equation 1 has been established, a measurement matrix $\mathbf{M}$ that takes advantage of the model needs to be devised. We chose to implement the simple optimization algorithm proposed in [10]:

$\mathbf{M} \in \mathbf{R}^{n_s \times n}$ is initialized as random binary matrix with precisely one 1 on each row. The algorithm then randomly replaces one row of $\mathbf{M}$ with a different random binary unit row vector. If the condition number $\kappa(\mathbf{MU})$ does not decrease, the change is reverted. This is repeated until convergence or a maximum number of steps is reached (*cf.* algorithm 1). For ABTFs, the condition number $\kappa((\mathbf{1}_{p^2} \otimes \mathbf{M})\mathbf{U})$ is tested instead.

The intuition behind this choice is that the condition number $\kappa(\mathbf{MU})$ is an indicator of how robustly $\mathbf{MU}$ can be inverted; *i.e.* of how well coefficients $\mathbf{C_B}$ as in equation 2 can be found.

In its present form the algorithm is free to choose whatever pairs of incoming and outgoing light directions lead to well-conditioned linear system. This approach suits best the gonioreflectometer setup, where all such pairs have equal costs. The algorithm can easily be modified to take the parallelism of camera dome setups into account.

While undersampling could be used in the kaleidoscope setup as well, we argue it is more beneficial to use the proposed method in order to increase the kaleidoscope's limited angular resolution.

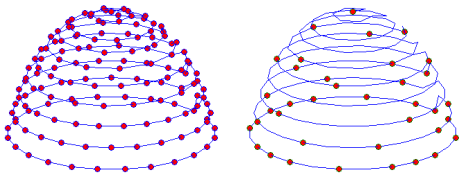
Algorithm 1 Generation of a measurement matrix.

Input: desired number n_s of samples
Output: optimized measurement matrix $\mathbf{M} \in \mathbf{R}^{n_s \times n}$

```
M ← random binary with exactly one 1 per row
while not converged do
    M' ← M
    r ← random binary row vector with ||r||_0 = 1
    random row of M ← r
    if κ(M'U) < κ(MU) then
        M ← M'
    end if
end while
return M
```

4 RESULTS

For our experiments, we used an existing database of high-quality measured BTFs. The measurement device used to create the database is a camera dome with 151 cameras, the flashes of which are used as light sources, resulting in an angular resolution of 151×151 (*cf.* figure 3a). The rectified textures have a spatial resolution of 512×512 pixels and correspond to a part of the sample approximately $4\text{ cm} \times 4\text{ cm}$ in size. The database consists of 14 semantic classes with 12 materials each. We selected the classes *carpet*, *cloth*, *gravel*, *leather*, *metal*, *stone*, *wall tile*, *wallpaper* and *wood*, which exhibit significant inter- and intraclass variance. We used 11 materials per class for fitting the linear models and the remaining material per class for the purpose of validation.



(a) (b)

Figure 3: (a) Sketch of the acquisition setup. The red points correspond to both light sources and cameras. (b) Sketch of the 7 × 7 parabolic map sampling. The red points correspond to both light sources and cameras.

All computations have been performed using MATLAB 2011b under Windows XP on a machine with two Intel Xeon E5645 processors and 144 GB of RAM.

4.1 Model fitting

Computing at once a truncated SVD for either of the entire database, or even single clusters thereof, is prohibitive due to the computation time required. We therefore used *eigenspace merging* to compute the SVD hierarchically; *i.e.* we first use EM-PCA (*cf.* [14]) to obtain approximate truncated SVDs of the single BTFs and subsequently merge the resulting eigenspaces (*cf. e.g.* [6]). In order to further reduce computation times, we cropped the BTFs to a spatial extent of 128×128 texels. For the purpose of comparison, we fit linear models both to the ABRDF and the ABTF database.

4.1.1 ABRDFs

For a single one of our BTFs, 200 eigen-ABRDFs for the $\log(Y)$ channel and 100 eigen-ABRDFs for each the U/Y and the V/Y channel provide a very high reproduction quality. We merged the resulting eigenspaces first per cluster and then globally to obtain an ABRDF basis of 2048 eigen-ABRDFs. The entire process takes approximately 30 minutes per cluster, including disk I/O and the color space transformation, hence about 4.5 hours altogether.

Table 1 shows the relative projection errors ε that occur when projecting the $\log(Y)$ channel of the test material's BTF onto the corresponding bases for various numbers of basis ABRDFs; *i.e.*

$$\varepsilon = \frac{\|\mathbf{U}(\mathbf{U}'\mathbf{B}) - \mathbf{B}\|_{\mathrm{F}}}{\|\mathbf{B}\|_{\mathrm{F}}}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. For comparison, we include the relative projection errors for the fully measured BTFs after FMF-compression retaining 128 eigen-ABRDFs, a number suitable for high-quality real-time rendering. Typically, 1024 basis ABRDFs are sufficient to achieve good projection results, which is the lower limit on the number of samples necessary for sparse reconstruction via equation 3.

4.1.2 ABTFs

For ABTFs, we computed bases per cluster. Following the argument in [17], we used a spatial ABTF size of 3×3. For performance reasons, we selected ABTFs maximally without overlap, resulting in a database of 1764 ABTFs per material and 19404 ABTFs per cluster. We again first computed bases per-material, retaining 200 eigen-ABTFs for the $\log(Y)$ channel and 100 eigen-ABTFs for each the U/Y and the V/Y channel, and then merged the resulting eigenspaces per cluster. This process takes approximately 2 hours per cluster, or 18 hours in total.

Table 1 shows relative projection errors (*cf.* section 4.1.1) for the $\log(Y)$ channel for 1024 and 2048 basis ABTFs in comparison with errors for reconstructions from ABRDF-wise projections. The projections themselves were produced by collecting all possible 3 × 3 ABTFs from the test BTF and projecting them onto the appropriate cluster's basis. BTFs are obtained from this representation by computing the reconstruction and blending the patches, all texels weighted equally. Typically, 2048 basis ABTFs provide almost as good projections results as 1024 basis ABRDFs.

4.2 Reconstruction

Figure 4 shows renderings of BTFs reconstructed with the proposed method, table 2 the corresponding relative reconstruction errors

$$\varepsilon = \frac{\|\mathbf{U}((\mathbf{M}\mathbf{U})^{\dagger}\widetilde{\mathbf{B}}) - \mathbf{B}\|_{\mathrm{F}}}{\|\mathbf{B}\|_{\mathrm{F}}}.$$

For comparison, we include renderings of the FMF-compressed original fully measured BTFs and their sparse reconstructions from ABRDF-wise linear models, along with the relative projection errors, which constitute lower limits for the relative reconstruction errors. BTFs were produced from ABTF-wise sparse reconstructions as described in section 4.1.2. We used two different sampling strategies: a 7 × 7 parabolic map mapped to the closest light and camera positions of the acquisition setup's full sampling, which may be considered a vague approximation of a kaleidoscope's sampling (*cf.* figure 3b), and optimized samplings with the same number of samples produced by algorithm 1. Both samplings consist of 1369 samples in total, or 6 % of the original 22801 samples.

4.2.1 ABRDFs

As predicted in section 3.2, ABRDF-wise reconstruction produces acceptable results only for materials with simple surface structure and reflectance – here: *stone* and *wood* – and even then only with the optimized sampling. *Leather* and *metal* already exhibit annoying artifacts; the results for even more complex materials are unsuitable for any practical purpose.

		# basis ABRDFs					# basis ABTFs	
class	FMF	128	256	512	1024	2048	1024	2048
carpet	5.4	6.2	5.6	5.1	4.6	4.0	5.3	5.1
cloth	2.7	4.0	3.2	2.6	2.1	1.7	2.9	2.7
gravel	5.0	8.0	6.7	5.6	4.7	4.0	5.4	4.9
leather	1.5	2.9	2.4	2.0	1.6	1.3	1.9	1.8
metal	1.0	3.2	2.5	2.1	1.8	1.6	2.7	2.2
stone	0.6	3.1	2.5	2.0	1.5	1.2	1.8	1.5
wall tile	0.4	5.6	4.7	3.8	2.9	2.1	2.2	1.7
wallpaper	2.7	5.2	4.4	3.6	3.0	2.4	3.4	3.1
wood	0.8	2.3	1.9	1.6	1.3	1.0	1.2	1.1

Table 1: $\log(Y)$ channel relative L^2 projection errors in percent for various numbers of basis ABRDFs. FMF: Projection onto per-material ABRDF basis with 128 eigen-ABRDFs.

	ABRDF			ABTF		
class	proj	pmap7	optimized	proj	pmap7	optimized
carpet	4.6	27.6	11.4	5.1	7.3	6.3
cloth	2.1	11.1	5.4	2.7	3.6	3.2
gravel	4.7	26.7	11.8	4.9	8.0	7.0
leather	1.6	8.0	4.0	1.8	2.3	2.1
metal	1.8	8.8	4.6	2.2	3.8	3.3
stone	1.5	6.3	3.4	1.5	2.3	2.3
wall tile	2.9	10.0	6.3	1.7	6.9	6.8
wallpaper	3.0	14.8	7.0	3.1	4.2	4.2
wood	1.3	6.3	2.9	1.1	1.6	1.5

Table 2: $\log(Y)$ channel relative L^2 reconstruction errors per cluster in percent. proj: Projection onto common basis. pmap7: Results for reconstruction from parabolic map sampling. optimized: Results for reconstruction from optimized sampling.

4.2.2 ABTFs

In contrast, even the non-optimized sampling is sufficient to produce convincing reconstructions of moderately complex materials using ABTF models. Where it is not, the optimized sampling often helps; only *gravel* and *wallpaper* exhibit perceivable artificats. The highlight of *wall tile* is not quite as sharp as it should be, and there are some artifacts in the highlight of *metal* visible mostly in the corresponding amplified error image (*cf.* figure 4f).

4.3 Limitations

While the proposed algorithm performs well in many situations, it has a number of limitations:

Most notably, it relies on the availability of a database of fully measured BTFs. Depending on the materials to be measured, that database must be quite encompassing; however, if *e.g.* only leathers are going to be measured, then a small database of a few measured leather BTFs might already be sufficient.

Without any regularization the lowest possible number of samples is precisely the number of basis ABTFs divided by the patch size. Typically, a greater number is necessary for robust results.

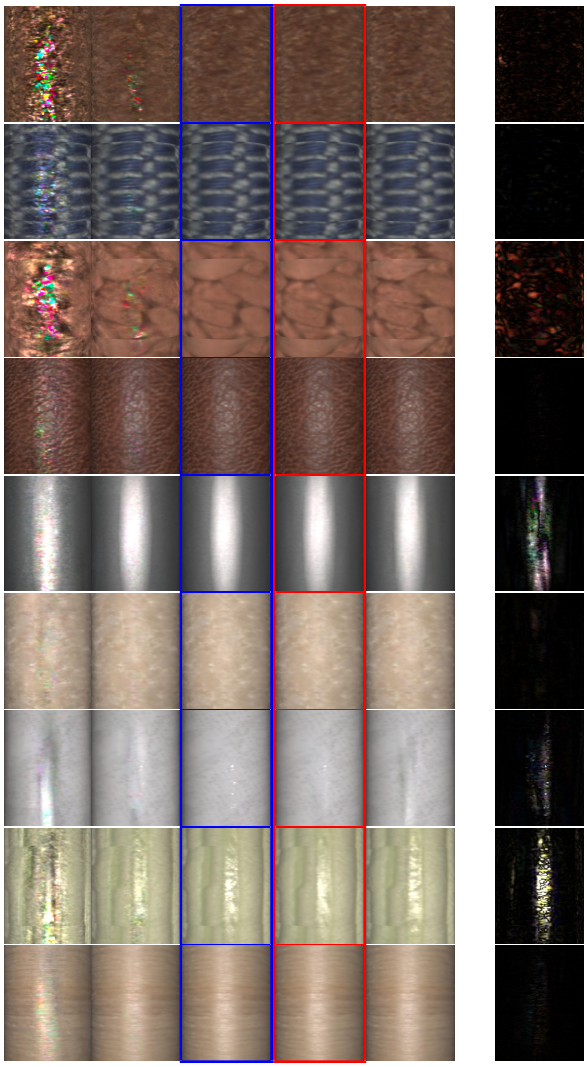
Reconstructions of material BTFs with highly complex surface structure may still suffer from artifacts visible in common lighting scenarios. It is not clear whether larger patch sizes could mitigate this. Even if so, this would likely lead to an undesirable significant increase of computation times and memory consumption.

For the same reason, the algorithm is constrained to moderate sampling rates. It would also be difficult to bootstrap a sufficiently large and heterogenous BTF database with substantially higher sampling rates.

5 CONCLUSION

We demonstrated the general possibility of efficient sparse acquisition of BTFs for a wide range of materials, provided a database of fully measured optically similar materials is available.

It would be interesting to investigate whether our results could be improved further. A possible approach is to further improve the linear bases, for instance by feature-aligning the ABTFs prior to fitting the models. It is also unclear how suitable our manual clustering of the database actually is. Automatic methods might be able to find a better optimization, possibly even consisting of fewer classes.



(a) (b) (c) (d) (e) (f)

Figure 4: Renderings of reconstruction results.
(a) ABRDF-based reconstructions from 7 × 7 parabolic map sampling (1369 samples).
(b) ABRDF-based reconstructions from optimized sampling (1369 samples).
(c) FMF-compressed ground truth (22801 samples).
(d) ABTF-based reconstructions from optimized sampling (1369 samples).
(e) ABTF-based reconstructions from 7 × 7 parabolic map sampling (1369 samples).
(f) 20 × absolute differences between (c) and (d).

Moreover, our linear models might also be useful for purposes other than sparse reconstruction; for instance, it might be possible to use them to leverage the quality of BTF measurements produced with consumer-grade hardware, or under conditions less controlled than in the discussed setups.

Although the improvement in measurement cost is significant, the amount of samples needed still leaves room for further improvement. Depending on the material, it should not be impossible — at least given a rough estimate of the material's surface structure — to obtain satisfactory reconstruction results from less than 100 images. Both our experiments and compressed sensing theory suggest, however, that this barrier cannot be broken merely using unregularized linear methods. It thus seems worthwhile to investigate non-linear methods such as manifold learning or texture synthesis.

6 ACKNOWLEDGEMENTS

This work was funded by the X-Rite graduate school on Digital Material Appearance.

7 REFERENCES

[1] Kristin J. Dana, Bram van Ginneken, Shree K. Nayar, and Jan J. Koenderink. Reflectance and texture of real-world surfaces. *ACM Trans. Graph.*, 18(1):1–34, January 1999.

[2] Yue Dong, Jiaping Wang, Xin Tong, John Snyder, Yanxiang Lan, Moshe Ben-Ezra, and Baining Guo. Manifold bootstrapping for SVBRDF capture. *ACM Trans. Graph.*, 29(4):98:1–98:10, July 2010.

[3] David L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52:1289–1306, 2006.

[4] Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936.

[5] Jiří Filip, Michael J. Chantler, Patrick R. Green, and Michal Haindl. A psychophysically validated metric for bidirectional texture data reduction. *ACM Trans. Graph.*, 27(5):138:1–138:11, December 2008.

[6] Peter Hall, David Marshall, and Ralph Martin. Merging and splitting eigenspace models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(9):1042–1049, September 2000.

[7] Jefferson Y. Han and Ken Perlin. Measuring bidirectional texture reflectance with a kaleidoscope. *ACM Trans. Graph.*, 22(3):741–748, July 2003.

[8] Kshitij Marwah, Gordon Wetzstein, Yosuke Bando, and Ramesh Raskar. Compressive light field photography using overcomplete dictionaries and optimized projections. *ACM Trans. Graph.*, 32(4):46:1–46:12, July 2013.

[9] Wojciech Matusik, Hanspeter Pfister, Matt Brand, and Leonard McMillan. A data-driven reflectance model. *ACM Trans. Graph.*, 22(3):759–769, July 2003.

[10] Wojciech Matusik, Hanspeter Pfister, Matthew Brand, and Leonard McMillan. Efficient isotropic BRDF measurement. In *Proceedings of the 14th Eurographics Workshop on Rendering*, EGRW

'03, pages 241–247, Aire-la-Ville, Switzerland, Switzerland, 2003. Eurographics Association.

[11] Gero Müller, Gerhard H. Bendels, and Reinhard Klein. Rapid synchronous acquisition of geometry and BTF for cultural heritage artefacts. In *The 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST)*, pages 13–20. Eurographics Association, Eurographics Association, November 2005.

[12] Gero Müller, Jan Meseth, and Reinhard Klein. Compression and real-time rendering of measured BTFs using local PCA. In *Vision, Modeling and Visualisation 2003*, pages 271–280. Akademische Verlagsgesellschaft Aka GmbH, Berlin, November 2003.

[13] Pieter Peers, Dhruv K. Mahajan, Bruce Lamond, Abhijeet Ghosh, Wojciech Matusik, Ravi Ramamoorthi, and Paul Debevec. Compressive light transport sensing. *ACM Trans. Graph.*, 28(1):3:1–3:18, February 2009.

[14] Sam Roweis. EM algorithms for PCA and SPCA. In *Proceedings of the 1997 Conference on Advances in Neural Information Processing Systems 10*, NIPS '97, pages 626–632, Cambridge, MA, USA, 1998. MIT Press.

[15] Christopher Schwartz, Ralf Sarlette, Michael Weinmann, and Reinhard Klein. DOME II: A parallelized BTF acquisition system. In *Eurographics Workshop on Material Appearance Modeling: Issues and Acquisition*, pages 25–31. Eurographics Association, June 2013.

[16] Christopher Schwartz, Ralf Sarlette, Michael Weinmann, Martin Rump, and Reinhard Klein. Design and implementation of practical bidirectional texture function measurement devices focusing on the developments at the university of bonn. *Sensors*, 14(5), April 2014.

[17] Manik Varma and Andrew Zisserman. A statistical approach to material classification using image patch exemplars. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(11):2032–2047, November 2009.

[18] Tien-Tsin Wong, Pheng-Ann Heng, Siu-Hang Or, and Wai-Yin Ng. Image-based rendering with controllable illumination. In *Proceedings of the Eurographics Workshop on Rendering Techniques '97*, pages 13–22, London, UK, UK, 1997. Springer-Verlag.

APPENDIX B

Multiplexed acquisition of bidirectional texture functions for materials

Best Student Paper Award

Multiplexed Acquisition of Bidirectional Texture Functions for Materials

Dennis den Brok, Heinz C. Steinhausen, Matthias B. Hullin and Reinhard Klein

Institut für Informatik II, Universität Bonn, Germany

ABSTRACT

The bidirectional texture function (BTF) has proven a valuable model for the representation of complex spatially-varying material reflectance. Its image-based nature, however, makes material BTFs extremely cumbersome to acquire: in order to adequately sample high-frequency details, many thousands of images of a given material as seen and lit from different directions have to be obtained. Additionally, long exposure times are required to account for the wide dynamic range exhibited by the reflectance of many real-world materials.

We propose to significantly reduce the required exposure times by using illumination patterns instead of single light sources ("multiplexed illumination"). A BTF can then be produced by solving an appropriate linear system, exploiting the linearity of the superposition of light. Where necessary, we deal with signal-dependent noise by using a simple linear model derived from an existing database of material BTFs as a prior. We demonstrate the feasibility of our method for a number of real-world materials in a camera dome scenario.

Keywords: digital material appearance, bidirectional texture functions, illumination multiplexing

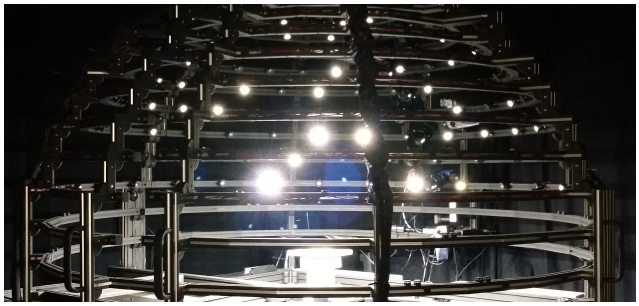


Figure 1. Our acquisition setup displaying an S-matrix pattern.

1. INTRODUCTION

Analytical or physically-based reflectance models such as the bidirectional reflectance distribution function (BRDF) and its spatially varying sibling (SVBRDF) generally fail to accurately represent the noticeable non-local effects such as interreflections and self-shadowing that can be observed on many common real-world materials. The image-based *bidirectional texture function* (BTF) accounts for these effects, even in real-time applications. This image-based nature, however, makes material BTFs extremely cumbersome to acquire: in order to adequately sample high-frequency details, many thousands of images of a given material with different lighting and viewing positions have to be obtained. Parallel setups such as camera domes and kaleidoscopes significantly accelerate this process, but they still suffer from long exposure times for high dynamic range (HDR) capture and the sheer number of samples required.

We propose to significantly reduce the required exposure times by using *multiplexed illumination*: instead of taking images for each light source seperately, we use patterns consisting of many light sources at once, greatly increasing the amount of light on the material sample. Exploiting the linearity of the superposition of light, it is then possible to reconstruct the corresponding single-light images by solving an appropriate linear system. In the presence of signal-dependent noise such as photon noise, this process may yield very noisy images, in particular if the brightness of the scene varies greatly.

Measuring, Modeling, and Reproducing Material Appearance 2015, edited by Maria V. Ortiz Segovia, Philipp Urban, Francisco H. Imai, Proc. of SPIE-IS&T Vol. 9398, 93980F · © 2015 SPIE-IS&T · CCC code: 0277-786X/15/$18 · doi: 10.1117/12.2078396

We demonstrate that for rather simple materials, this effect is not noticeable in practice, and for moderately complex materials where it is, it is possible to denoise the resulting BTFs by using a simple linear model derived from an existing database of material BTFs as a prior.

We evaluate the proposed method on a number of real-world materials in a camera dome scenario and show that it is possible to reduce acquisition times by about 75–95%.

2. RELATED WORK

2.1 Plenoptic multiplexing

An in-depth theoretical background on plenoptic multiplexing in general has been given by Hartwit and Sloane.[1] They prove or conjecture the optimality of Hadamard patterns and their binary derivates with respect to several measures. They also cover the various noise sources in optical systems and how they influence demultiplexing, albeit quite abstractly so.

As far as the authors are aware, Wenger *et al.*[2] were among the first to investigate multiplexing for the purpose of capturing some kind of appearance; in this case, time-varying light fields of human faces. However, they observed an intolerable amount of noise when using Hadamard patterns that they were unable to reduce to a tolerable level through simple filtering.

In contrast, Schechner *et al.*[3] deal with the special case of illumination multiplexing using digital photo cameras and various types of light sources. They explain in great detail the implications of this scenario regarding noise and derive a formula useful for judging whether multiplexing is beneficial in a specific situation, taking into account the setup's relevant intrinsic parameters.

Ratner *et al.*[4] provide an optimization method which produces illumination patterns that take the noise characteristics of a given setup into account. They assume a one-dimensional affine noise model and a nearly diffuse scene, assumptions generally violated in BTF acquisition. Furthermore, they show that in the presence of overexposure it is preferable to reduce the number of light sources instead of shutter times.

Mitra *et al.*[5] take this even further and compute illumination patterns using an optimization based on image priors. In contrast to previous methods, their method is able to handle large amounts of light, but it still relies on the assumption of a one-dimensional affine noise model. They show how to extend their method to low-resolution light fields, which is, however, computationally very expensive and therefore likely prohibitive in the case of BTFs.

2.2 Sparse acquisition

Sparse acquisition is related with the proposed method not directly in terms of methodology, but in that its goal is faster acquisition, and to that end usually some kind of model is used as a prior as well. We thus briefly review a number of articles on sparse acquisition.

Matusik *et al.*[6] perform *singular value decomposition* (SVD) on a database of 100 measured BRDFs of a wide range of isotropic materials to obtain a linear model.

The same authors[7] introduce two methods for sparse reconstruction of isotropic BRDFs: The first method is based on a wavelet analysis of their BRDF database. A set of basis wavelets termed *common wavelet basis* is determined and used to reconstruct previously unseen BRDFs with approximately 1.5 million samples from approximately 70000 measurements. The second method uses the entire BRDF database itself as a linear model for reconstruction of fully measured BRDFs from as little as 800 out of the original approximately 1.5 million samples, at the cost of slightly increased reconstruction errors and the required availability of the BRDF database. Samples are chosen using a simple optimization algorithm such that the linear system to be solved for reconstruction is well-conditioned. They do not investigate how well their methods generalize to more complex reflectance such as anisotropic BRDFs.

Filip *et al.*[8] propose a vector quantization of BTFs for the purpose of compression, guided by a psychophysically validated metric. They conclude that as little as $10 - 35$ % of the original textures are sufficient to maintain the same visual appearance in renderings. It would be interesting to investigate whether there is a common quantization for *all* materials, and if so, whether it could be used for sparse acquisition. A large user study would be needed in order to adapt the metric to a bigger BTF database.

Peers *et al.*[9] introduce *compressed sensing*[10] to the acquisition of *reflectance fields*, assuming both 2D *outgoing* (here: fixed viewing direction) and *incident light fields*. Their algorithm uses a hierarchical, multi-resolution Haar wavelet basis, taking spatial coherence into account. It is not clear how to extend this approach to a multi-view setup. Common BTF acquisition setups only have a very limited number of light sources, where the advantage of compressed sensing might be negligible. Due to these light sources' brightness, we expect shot noise to become a problem.

Dong *et al.*[11] reconstruct a material's SVBRDF from a sparse measurement using a manifold constructed from analytical BRDFs fit to fully measured BRDFs of manually selected representative points on the material's surface. The algorithm is unlikely to scale to BTFs because of the typically much higher intrinsic dimensionality of the manifold of per-texel reflectance distribution functions. A generalization to previously unseen materials is not obvious, albeit conceivable.

Marwah *et al.*[12] use sparsity-based methods related to compressed sensing in order to sparsely acquire 4D light fields with an angular resolution of 5×5. They compute a dictionary of what they call *light field atoms* – 11×11 spatial light field patches which allow for a sparse representation of natural light fields. Such a dictionary does generally not exist in the case of BTFs; their dimensionality is likely too high.

Inspired by the work of Matusik *et al.* we[13] demonstrated that there exists linear models that lend themselves to sparse reconstruction of BTFs. We proposed to fit linear models to small BTF patches in order to account for non-local effects, with seperate models per (manually chosen) classes of materials to constrain the dimensionality of that data. Straight-forward models fit to per-texel reflectance distribution functions proved insufficient for our purpose.

3. BACKGROUND

3.1 Bidirectional texture functions

BTFs have been introduced by Dana *et al.*[15] as an image-based approach to spatially varying appearance. Like SVBRDFs, they are 6-dimensional functions of the form

$$\mathcal{B}(\mathbf{x}, \omega_i, \omega_o),$$

where $\omega_{i,o} \in \mathbf{R}^2$ are the incoming and outgoing light directions, respectively, and $\mathbf{x} \in \mathbf{R}^2$ is the position on a parameterized surface V. In the case of material BTFs, V is typically flat; it does not need to coincide with the material's actual surface geometry. It is generally assumed that light sources are directional and have the same spectrum. In particular, effects such as phosphorescene, fluorescence and subsurface scattering cannot be captured accurately.

The fundamental difference from SVBDRFs is that the function $\mathcal{B}(\mathbf{x}, -)$ need not be BRDF-valued: the corresponding per-texel reflectance function does not need to adhere to Helmholtz reciprocity and conservation of energy and is, therefore, capable of capturing non-local effects such as interreflections and self-shadowing. Moreover, because V does not necessarily coincide with the actual surface, the per-texel reflectance functions may also describe parallax effects. For these reasons, the term *apparent BRDF* (ABRDF) has been suggested by Wong *et al.*[16] for this kind of functions. Conversely, the values of the function $\mathcal{B}(-, \omega_i, \omega_o)$ are just 2D textures corresponding to specific pairs of incoming and outgoing light directions.

During measurement, a finite discretization of the measured material's BTF $\mathcal{B}$ is obtained. After rectification of the acquired images, the discrete BTF has a natural representation as a matrix $\mathbf{B} \in \mathbf{R}^{n \times m}$ with the columns representing the m discrete ABRDFs, each entry corresponding to some pair (ω_i, ω_o) of incoming and outgoing angle, and the rows representing the n rectified textures (*cf.* figure 2).

3.2 BTF acquisition

Several setups for the acquisition of BTFs have been proposed. We briefly review the most prominent paradigms, as our method benefits all of them to a greater or lesser extent. An up-to-date, in-depth overview is given in the recent survey by Schwartz *et al.*[17]

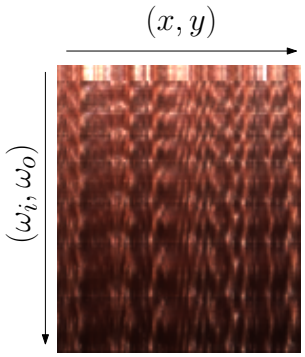


Figure 2. Representation of a discretized BTF as a matrix.

3.2.1 Gonioreflectometer

In what is historically the first BTF acquisition setup, proposed by Dana *et al.*,[15] the material sample is placed on a turntable, and a camera and a light source held by robot arms are moved across the hemisphere above the sample to capture images of the sample under different lighting and viewing conditions. The gonioreflectometer is very flexible in terms of possible samplings of the hemisphere, but measurement times are excessive – on the order of weeks for a moderate angular resolution – due to the little amounts of light sources and sensors and the movable parts' low speeds.

Due to the absence of multiple light sources, our method does not readily apply to traditional setups. If the gonioreflectometer's total flexibility is not a necessity for the envisioned scenario, additional light sources can be added.

3.2.2 Kaleidoscope

Han *et al.*[18] introduced an intriguing parallel setup: The sample is placed underneath a tapered kaleidoscope, lit and captured from a projector and a camera placed at the other end, which allows for a number of lighting and viewing conditions to be measured in a single camera shot. By appropriately arranging the mirrors, the angular and spatial resolution can be adjusted; however, both are typically rather low, and increasing one leads to a decrease of the other, so there is always a tradeoff to be made.

3.2.3 Camera domes

Camera domes as proposed, for instance, by Müller *et al.*[19] and Schwartz *et al.*[20] ideally provide a highly parallel means to acquire BTFs: A number of cameras is spread across the hemisphere above the sample holder. Their flashes or separate LEDs are used as light sources. Parallelism may be traded for fewer cameras and lower cost by placing the sample on a turntable in order to achieve a similarly dense sampling of the hemisphere. Due to the number of cameras, data transfer times become a new bottleneck.

In all of the above setups, it is usually necessary to capture the same scene several times with different shutter times in order to obtain HDR data.

3.3 Illumination multiplexing

Multiplexed measurement is based on the observation that by using appropriate patterns, the amount of *signal-independent* noise in the demultiplexed measurements will be lower than without multiplexing. In the case of imaging systems, the resulting linear system to be solved is

$$\mathbf{M} \cdot \mathbf{I}_{\text{single}} = \mathbf{I}_{\text{multiplexed}}$$

where $\mathbf{M} \in \mathbf{Z}^{n \times n}$, $\mathbf{I}_{\text{single}} \in \mathbf{R}^{n \times (w \cdot h)}$ with each row a $w \times h$ image of the scene lit by an individual light source, and $\mathbf{I}_{\text{multiplexed}} \in \mathbf{R}^{n \times (w \cdot h)}$ where each row is a $w \times h$ image of the scene lit by an individual illumination pattern.

Typically, $\mathbf{M}$ is supposed to be binary, corresponding to light sources of equal power that can only be on or off. It has been shown[1] that binary illumination patterns given by a certain kind of matrices called *S-matrices*

are at least very close to minimizing the average mean square error, and that the associated signal-to-signal-independent-noise ratio is increased by a factor of $\sqrt{n}/2$ for large n. It has, however, also been shown that the presence of signal-dependent noise such as photon noise strongly counteracts this advantage.[13]

There are several ways to produce S-matrices $\mathbf{M} \in \mathbf{N}^{n \times n}$ of a given order $n \in \mathbf{N}$;[1] we chose to implement the most straight-forward one, the so-called *quadratic residue construction*: Let $n = 4p + 3$, n, $p \in \mathbf{N}$ prime, be the desired order, and Q the set of quadratic residues of $0, \ldots, \frac{n-1}{2}$ modulo n. Then

$$m_{1j} = \begin{cases} 1, & j \in Q \\ 0, & \text{else} \end{cases}$$

The remaining $n - 1$ rows of $\mathbf{M}$ are obtained by cyclically shifting the preceding row to the left by one.

Clearly, the number of light sources in a given acquisition setup need not necessarily be prime of the required form. In that case, one can choose a suitable higher order and truncate the resulting S-matrix such that the number of columns equals the number of light sources, and solve the resulting linear system in the least-squares sense.[1]

4. MULTIPLEXED BTF ACQUISITION

4.1 Experimental setup

We chose 12 materials from 3 classes – *cloth*, *leather*, and *wood* – each. Per class, we picked 4 materials for the purpose of verifying the proposed method which we measured both conventionally and using the proposed method.

The measurement device we used in our experiments is a camera dome with 11 industrial-grade cameras with a maximum frame-rate of 8 Hz and 198 LED light sources, 10 of which are placed opposing the lower-most 10 cameras to produce direct reflections. The quadratic residue construction requires the S-matrix order to be prime of the form $4p + 3$, p prime. The smallest S-matrix order of that form greater or equal the number of LEDs in our setup is $n = 199$, which we chose for our experiments.

In order to speed up the acquisition of the ground truth data, the single-light measurements were obtained using a camera gain of 10 dB, while the multiplexed measurements were obtained with a camera gain of 0 dB. Additionally, we measured 4 small patches (approximately 1 cm $\times$ 1 cm) of different materials of the same class at once. We argue that, even in practive, this is a reasonable trade-off to be made: If required, larger BTFs for each single sample can be produced quickly by measuring them again using the proposed method, or by extrapolation as described *e.g.* by Steinhausen *et al.*[21]

The materials are placed on a turntable, which during measurement is rotated 12 times by 30°, in order to achieve a dense sampling of viewing directions. For each turntable position, the cameras take pictures of the material lit by each of the LEDs separately using several exposure times determined manually before the actual acquisition. The test materials were additionally measured using S-matrix patterns of order $n = 199$ before proceeding to the next turntable position.

After measurement, we combine the low dynamic range (LDR) raw images into high dynamic range (HDR) images and subsequently demosaic and rectify them. We use LED calibration data in order to account for variations in the LEDs' spectra. For further details on the various post-processing steps, *cf.* Schwartz *et al.*[17]

Finally, the resulting images are arranged and stored as matrices, which in the case of multiplexed measurements are then demultiplexed. For rendering, the BTFs are furthermore resampled in the angular domain, such that the light and view hemispheres are the same for each texel, and compressed using truncated SVD.

	$\varepsilon_{\text{demult}}$	$\varepsilon_{\text{denoised}}$
cloth 1 – 4	13.1, 10.7, 9.4, 9.7 (⌀ 10.2)	6.2, 5.6, 5.3, 5.1 (⌀ 5.6)
leather 1 – 4	13.1, 15.7, 17.1, 19.9 (⌀ 16.5)	6.5, 6.2, 6.9, 7.1 (⌀ 6.7)
wood 1 – 4	7.0, 7.3, 8.5, 7.0 (⌀ 7.5)	3.5, 3.6, 4.0, 3.6 (⌀ 3.7)

Table 1. Comparison of log(Y) relative errors [%], demultiplexed and denoised.

Figure 3. Renderings of Wood 1. Left: demultiplexed. Center: ground truth. Right: denoised.

4.2 Linear models and denoising

For the purpose of denoising, we fit linear models seperately to the per-color channel ABRDFs of each of the three classes, converted to log(Y) U/Y V/Y color space, without the test materials. The fitting is done using a truncated SVD

$$\mathbf{D}_{\text{class}} \approx \mathbf{U}\Sigma\mathbf{V}^t,$$

which is known by the Eckhart-Young theorem to be optimal in terms of L^2 error,[22] where $\mathbf{D}_{\text{class}}$ is a matrix with the ABRDFs of the BTFs of class *class* as columns. We have shown[13] that these model generalize to materials $\mathbf{B}$ that do not belong to the particular database; *i.e.*

$$\min_{\mathbf{C_B}} ||\mathbf{UC_B} - \mathbf{B}|| < \varepsilon.$$

For our experiments, we used 768 basis vectors for log(Y) and 128 for both U/Y and V/Y, which, according to our previous experiments,[13] is sufficient for the materials classes considered in the present article. Under the assumption that demultiplexing noise cannot be represented in these bases, denoising can now be performed by simply projecting a demultiplexed BTF onto the corresponding basis:

$$\mathbf{B}_{\text{denoised}} = \mathbf{U} \cdot (\mathbf{U}^t \cdot \mathbf{B}_{\text{demultiplexed}}).$$

5. RESULTS

Table 1 shows relative errors

$$\varepsilon_{\{\text{demult,denoised}\}} = \frac{\|\mathbf{B}_{\{\text{demult,denoised}\}} - \mathbf{B}_{\text{reference}}\|_{\mathrm{F}}}{\|\mathbf{B}_{\text{reference}}\|_{\mathrm{F}}}$$

of the demultiplexed and denoised demultiplexed BTFs, respectively. Errors are computed on the log(Y) color channel to account for human perception.

In the case of wood, the relative errors are relatively small, and indeed there are no obvious differences between the renderings of ground truth, demultiplexed and denoised BTF, even though the relative error is approximately halved by denoising (*cf.* figure 3).

In contrast, Cloth 1 and Leather 2 exhibit annoying artifacts for grazing viewing angles, as seen at the cylinder's borders. While the denoised BTF does not look completely identical to the ground truth, it looks much more plausible at the borders (*cf.* figures 4 and 6). For better comparison, figure 5 shows textures corresponding to a low camera position ($\theta = 75°$), extracted from Cloth 1's BTF. Note that these are rectified images; in an actual rendering, each pixel will be a weighted average of several texels.

Demultiplexed Leather 4, however, exhibits so much noise (yielding a relative error of almost 20%) that our denoising strategy breaks down as well (*cf.* figure 7), even though the relative error is more than halved.

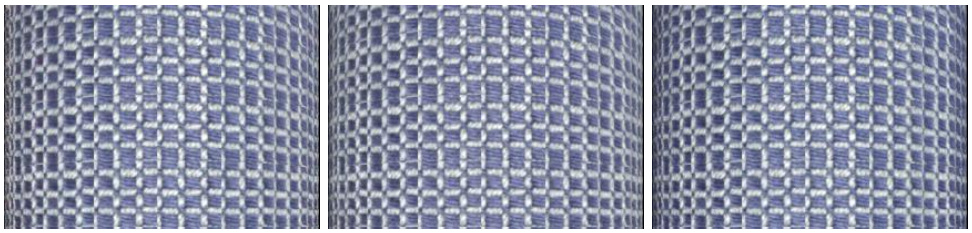


Figure 4. Renderings of Cloth 1. Left: demultiplexed. Center: ground truth. Right: denoised.

Figure 5. Low camera ($\theta = 75°$) textures extracted from Cloth 1 BTF. Left: demultiplexed. Center: ground truth. Right: denoised.

Figure 6. Renderings of Leather 2. Left: demultiplexed. Center: ground truth. Right: denoised.

Figure 7. Renderings of Leather 4. Left: demultiplexed. Center: ground truth. Right: denoised.

	single	multiplexed	rel. Δ
Cloth 1 – 4	10.8	1.8	−83%
Leather 1 – 4	23.6	1.2	−95%
Wood 1 – 4	4.4	1.2	−75%

Table 2. Comparison of acquisition times [h], single light vs. multiplexed.

As can be seen in table 2, the larger amount of light reaching the material samples on each particular image makes it possible to reduce total acquisition times by about 75–95%, even though the ground truth data was obtained with a much higher camera gain. In the case of Leather 1–4 and Wood 1–4, the minimum acquitision time achievable with our setup, caused by several bottlenecks, is reached.

6. CONCLUSION

We demonstrated that it is possibly to produce plausibly looking material BTFs from images taken under multiplexed illumination, possibly by using linear models derived from an existing database of material BTFs as a prior for denoising. The increased amount of light on the material sample can be used to considerably reduce the necessary shutter times, leading to total acquisition times reduced by 75–95%. As expected, we observed that this works best for materials with rather limited dynamic range.

It would be interesting to incorporate the method proposed by Mitra *et al.*,[5] which would, however, require both a suitable noise model for our camera dome setup and a modified optimization algorithm accounting for that model. Moreover, it could be possible to automatically detect particularly noisy textures within a demultiplexed BTF using adequate image statistics. These could then be regarded as missing values, and filled using *e.g.* our sparse reconstruction method.[13]

Finally, it is conceivable that a database to derive linear models for denoising from can be "bootstrapped", for instance by using smaller S-matrix orders, or by leaving out light sources that over-proportionally increase the BTF's dynamic range and measuring the corresponding images seperately.

7. ACKNOWLEDGEMENTS

This work was funded by the X-Rite graduate school on Digital Material Appearance. We thank Roland Ruiters and Martin Rump for helpful suggestions.

REFERENCES

[1] Harwit, M. and Sloane, N., [*Hadamard transform optics*], Academic Press (1979).

[2] Wenger, A., Gardner, A., Tchou, C., Unger, J., Hawkins, T., and Debevec, P., "Performance relighting and reflectance transformation with time-multiplexed illumination," *ACM Trans. Graph.* **24**, 756–764 (July 2005).

[3] Schechner, Y. Y., Nayar, S. K., and Belhumeur, P. N., "Multiplexing for optimal lighting," *IEEE Trans. Pattern Anal. Mach. Intell.* **29**, 1339–1354 (Aug. 2007).

[4] Ratner, N. and Schechner, Y., "Illumination multiplexing within fundamental limits," in [*Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*], 1–8 (June 2007).

[5] Mitra, K., Cossairt, O., and Veeraraghavan, A., "Can we beat hadamard multiplexing? data driven design and analysis for computational imaging systems," in [*Computational Photography (ICCP), 2014 IEEE International Conference on*], 1–9 (May 2014).

[6] Matusik, W., Pfister, H., Brand, M., and McMillan, L., "Efficient isotropic BRDF measurement," in [*Proceedings of the 14th Eurographics Workshop on Rendering*], *EGRW '03*, 241–247, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland (2003).

[7] Matusik, W., Pfister, H., Brand, M., and McMillan, L., "A data-driven reflectance model," *ACM Trans. Graph.* **22**, 759–769 (July 2003).

[8] Filip, J., Chantler, M. J., Green, P. R., and Haindl, M., "A psychophysically validated metric for bidirectional texture data reduction," *ACM Trans. Graph.* **27**, 138:1–138:11 (Dec. 2008).

[9] Peers, P., Mahajan, D. K., Lamond, B., Ghosh, A., Matusik, W., Ramamoorthi, R., and Debevec, P., "Compressive light transport sensing," *ACM Trans. Graph.* **28**, 3:1–3:18 (Feb. 2009).

[10] Donoho, D. L., "Compressed sensing," *IEEE Trans. Inform. Theory* **52**, 1289–1306 (2006).

[11] Dong, Y., Wang, J., Tong, X., Snyder, J., Lan, Y., Ben-Ezra, M., and Guo, B., "Manifold bootstrapping for SVBRDF capture," *ACM Trans. Graph.* **29**, 98:1–98:10 (July 2010).

[12] Marwah, K., Wetzstein, G., Bando, Y., and Raskar, R., "Compressive light field photography using overcomplete dictionaries and optimized projections," *ACM Trans. Graph.* **32**, 46:1–46:12 (July 2013).

[13] den Brok, D., Steinhausen, H. C., Hullin, M. B., and Klein, R., "Patch-based sparse reconstruction of material BTFs," *Journal of WSCG* **22**, 83–90 (June 2014).

[14] Filip, J., Vávra, R., and Krupicka, M., "Rapid material appearance acquisition using consumer hardware," *Sensors* **14**(10), 19785–19805 (2014).

[15] Dana, K. J., van Ginneken, B., Nayar, S. K., and Koenderink, J. J., "Reflectance and texture of real-world surfaces," *ACM Trans. Graph.* **18**, 1–34 (Jan. 1999).

[16] Wong, T.-T., Heng, P.-A., Or, S.-H., and Ng, W.-Y., "Image-based rendering with controllable illumination," in [*Proceedings of the Eurographics Workshop on Rendering Techniques '97*], 13–22, Springer-Verlag, London, UK, UK (1997).

[17] Schwartz, C., Sarlette, R., Weinmann, M., Rump, M., and Klein, R., "Design and implementation of practical bidirectional texture function measurement devices focusing on the developments at the university of bonn," *Sensors* **14** (Apr. 2014).

[18] Han, J. Y. and Perlin, K., "Measuring bidirectional texture reflectance with a kaleidoscope," *ACM Trans. Graph.* **22**, 741–748 (July 2003).

[19] Müller, G., Bendels, G. H., and Klein, R., "Rapid synchronous acquisition of geometry and BTF for cultural heritage artefacts," in [*The 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST)*], 13–20, Eurographics Association, Eurographics Association (Nov. 2005).

[20] Schwartz, C., Sarlette, R., Weinmann, M., and Klein, R., "DOME II: A parallelized BTF acquisition system," in [*Eurographics Workshop on Material Appearance Modeling: Issues and Acquisition*], 25–31, Eurographics Association (June 2013).

[21] Steinhausen, H. C., den Brok, D., Hullin, M. B., and Klein, R., "Acquiring bidirectional texture functions for large-scale material samples," *Journal of WSCG* **22**, 73–82 (June 2014).

[22] Eckart, C. and Young, G., "The approximation of one matrix by another of lower rank," *Psychometrika* **1**(3), 211–218 (1936).

APPENDIX C

Fast multiplexed acquisition of high-dynamic-range material appearance

Vision, Modeling, and Visualization (2015)
D. Bommes, T. Ritschel, and T. Schultz (Eds.)

Fast multiplexed acquisition of high-dynamic-range material appearance

D. den Brok,[1] H. C. Steinhausen[1] and R. Klein[1]

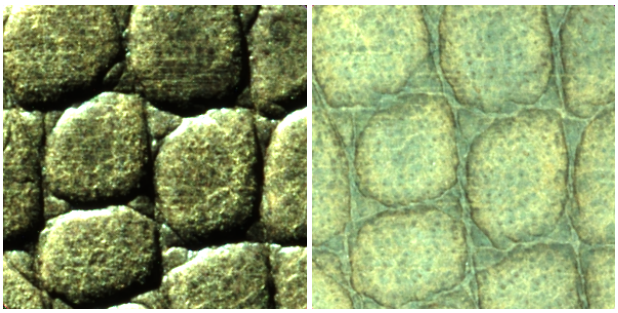


Figure 1: *Left: leather sample under single-light illumination. Right: same sample under multiplexed illumination.*

Abstract

There is tremendous demand for digital representations that allow for materials to be re-lit under as wide a range of illumination scenarios as possible. It is therefore desirable to capture the entire dynamic range of a material's appearance. This process can require excessive shutter times for many materials that reflect only small amounts of light for certain lighting and viewing directions, for instance in the presence of low albedo or self-shadowing. The problem is amplified in the case of image-based appearance models such as the bidirectional texture function (BTF), where possibly many thousands of images are required to accurately sample high-frequency details in the angular domain. We propose to capture material BTFs with their dynamic range compressed by multiplexed illumination. We demonstrate that the signal-dependent noise associated with demultiplexing can be mitigated by means of an existing database of low-noise material BTFs. Moreover, we investigate a method to quickly create a suitable database from scratch.

Categories and Subject Descriptors (according to ACM CCS): I.4.1 [Image Processing and Computer Vision]: Digitization and Image Capture—Reflectance

1. Introduction

In many applications it is desirable to be able to predict a given material's appearance under illumination that differs vastly from the controlled illumination that is typically employed when obtaining a material's digital representation. It is therefore important to capture as much of the dynamic range of material's reflectance as possible. The process of doing so, however, may require exposure series of many separate steps with shutter times that can range from only a few milliseconds up to minutes, for instance in the presence of low albedo or self-shadowing. In the case of image-based appearance models such as the *bidirectional texture function* (BTF), where the number of single images contributes strongly to the quality of reproductions in renderings, the resulting resource requirements can become excessive: Due to the many and possibly long shutter times, acquiring a single material's BTF may take from hours to even days. Moreover, a BTF typically comprises many thousands of (HDR) images, which during acquisition is multiplied by the number of (LDR) images per exposure series. This can easily lead to terabytes of data and increases the time spent on post-processing.

We propose to acquire material BTFs using *illumination multiplexing*: the material sample is illuminated not by each

© The Eurographics Association 2015.

single light source individually, but with a series of patterns of many light sources at once. That way, the dynamic range of a material's reflectance is compressed, because less shadows and highlights appear due to the many different simultaneous illumination angles, and the material is exposed to a larger amount of light. A BTF can thus be obtained with shorter exposure series and greatly reduced shutter times. Recovery of the desired images under single-light illumination amounts to solving an appropriate linear system, exploiting the linearity of the superposition of light. In the presence of signal-dependent noise, this *demultiplexing* process may yield intolerably noisy images, in particular if the material's dynamic range is wide.

We demonstrate that this effect can be mitigated by projecting the noisy BTF onto a linear subspace spanned by an existing database of traditionally measured material BTFs. In order to make this projection more robust, we heuristically identify possibly problematic images to be treated as missing values during projection.

As a suitable database may not always be available, we investigate a method of speeding up the process of creating one by using a combination of single-light and multiplexed illumination. We show that a database with material BTFs thus obtained performs not much worse in de-noising but is much quicker to create.

We evaluate our method on a number of real-world materials in a camera dome setup and show that it is possible to reduce acquisition times by about 75–95%, often down to our setup's physical limits, while maintaining a satisfactory reproduction quality.

2. Related work

2.1. Multiplexed illumination

Illumination multiplexing belongs to the wider field of plenoptic multiplexing. An extensive theoretical introduction to the topic has been provided by Hartwit and Sloane [HS79]. Essential for the present article is their proof of the optimality of Hadamard patterns and their binary siblings with respect to a number of measures. They also discuss the various noise sources in optical systems and their influence on demultiplexing on a high level.

Wenger et al. [WGT*05] propose multiplexing for the purpose of capturing time-varying light fields of human faces. However, they observed an intolerable amount of noise when using Hadamard patterns that they were unable to reduce to a tolerable level through simple filtering.

Schechner et al. [SNB07] provide a careful analysis of illumination multiplexing using digital photo cameras and various types of light sources, paying great attention to the possible kinds of noise. They arrive at a formula as a criterion whether multiplexing is beneficial given the setup's relevant intrinsic parameters.

Ratner et al. [RS07] demonstrate an optimization method which produces illumination patterns that take the noise characteristics of a given setup into account. They assume a one-dimensional affine noise model and a nearly diffuse scene, which does not lend itself well to BTF acquisition. Furthermore, they show that in the presence of overexposure it is preferable to reduce the number of light sources instead of shutter times.

Mitra et al. [MCV14] take this even further and compute illumination patterns using an optimization based on image priors. In contrast to previous methods, their method is able to handle large amounts of light, but it still relies on the assumption of a one-dimensional affine noise model. They show how to extend their method to low-resolution light fields, which is, however, computationally very expensive and therefore likely prohibitive in the case of BTFs.

As far as the authors are aware, den Brok et al. [dBSHK15] are the first to use multiplexed illumination in the context of BTF acquisition. They, too, use a database of previously acquired materials for the purpose of de-noising; they do, however, not provide a means to quickly bootstrap such a database, and we shall demonstrate that their de-noising scheme can be improved upon.

2.2. Sparse acquisition

We briefly review a number of articles on sparse acquisition, even though not directly related, because its purpose, too, is to reduce acquisition times.

Matusik et al. [MPBM03] demonstrate two sparse reconstruction methods for measured isotropic BRDFs: In the first method, they determine a set of basis wavelets they use to reconstruct previously unseen BRDFs with approximately 1.5 million samples from approximately half that amount of measurements. In the second, they reconstruct fully resolved measured BRDFs from 800 out of the original 1.5 million samples using their entire training BRDF database as a linear model. Drawbacks are slightly increased reconstruction errors and the required availability of the BRDF database. They do not investigate how well their methods generalize to more general appearance models.

Koudelka et al. [KMBK03] use single per-material linear models for apparent BRDFs for the purpose of BTF compression.

Peers et al. [PML*09] introduce *compressed sensing* [Don06] to the acquisition of *reflectance fields*, assuming both 2D *outgoing* (here: fixed viewing direction) and *incident light fields*. Their algorithm uses a hierarchical, multiresolution Haar wavelet basis that takes spatial coherence into account. It is unclear how to extend their approach to BTFs, where multiple viewing directions come into play, and the typically very limited number of light sources counteracts the advantage of compressed sensing. We expect shot noise to become a problem in this scenario as well.

© The Eurographics Association 2015.

Dong et al. [DWT*10] reconstruct a material's SVBRDF from a sparse measurement using a manifold constructed from analytical BRDFs fit to fully measured BRDFs of manually selected representative points on the material's surface. The algorithm is unlikely to scale to BTFs because of the typically much higher intrinsic dimensionality of the manifold of per-texel reflectance distribution functions. A generalization to previously unseen materials is not obvious, albeit conceivable.

Marwah et al. [MWBR13] use sparsity-based methods related to compressed sensing in order to sparsely acquire 4D light fields with an angular resolution of 5×5. They compute a dictionary of what they call *light field atoms* – 11×11 spatial light field patches which allow for a sparse representation of natural light fields. It is unlikely that such a dictionary exists in the case of BTFs, as their dimensionality is much higher.

Following Matusik et al. [MPBM03], den Brok et al. [dBSHK14] demonstrate that there exist linear models that lend themselves to sparse reconstruction of BTFs. In order to account for non-local effects, they propose to fit the models to small BTF patches, manually clustered by material class to constrain the dimensionality of that data. That way, they are able to reconstruct BTFs with the acquisition setup's full resolution from 6% of the total number of sample images.

Miandji et al. [MKU15] introduce a novel compressed sensing framework they demonstrate to work for 2D images, videos, and even 4D light fields. They use 2D patches from training data and show how to convert the problem of 2D sparse signal recovery to an equivalent problem in 1D. It would be interesting to investigate if their method could be lifted to 6D BTFs.

3. Background

3.1. Bidirectional texture functions

BTFs have been used in practice first by Dana et al. [DvGNK99]. Like spatially-varying BRDFs, they are 6-dimensional functions of the form

$$\mathcal{B}(\mathbf{x}, \omega_i, \omega_o),$$

where $\omega_{i,o} \in \mathbf{R}^2$ denote the directions of incoming and outgoing light, respectively, and $\mathbf{x} \in \mathbf{R}^2$ denotes the position on a parameterized surface V which, in the case of material BTFs, V is typically assumed flat; it does not need to coincide with the material's actual surface geometry. Light sources are usually assumed to be directional and have the same spectrum, which means that effects such as phosphorescene, fluorescence and subsurface scattering cannot be accounted for accurately.

Note that the values of the function $\mathcal{B}(\mathbf{x}, -)$ are not BRDFs in the strict sense: they typically do not adhere to Helmholtz reciprocity and conservation of energy and, in contrast to BRDFs, are therefore capable of capturing non-local effects such as interreflections and self-shadowing. When V does not coincide with the material's actual surface, they also describe parallax effects. For these reasons, the term *apparent BRDF* (ABRDF) has been established by Wong et al. [WHON97] for this kind of functions. The values of the function $\mathcal{B}(-, \omega_i, \omega_o)$, in contrast, are straightforward 2D textures corresponding to individual pairs of incoming and outgoing light directions.

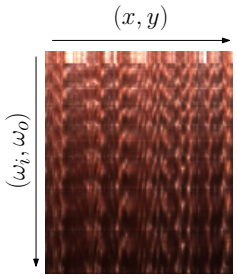


Figure 2: *Representation of a discretized BTF as a matrix.*

Discrete BTFs have a natural representation as a matrix $\mathbf{B} \in \mathbf{R}^{n \times m}$, where the columns represent the $w \times h$ discrete ABRDFs, each entry of which corresponds to a pair (ω_i, ω_o) of incoming and outgoing light direction, and the rows correspond to the n rectified 2D textures, where n is the number of distinct pairs of incoming and outgoing light directions (cf. Fig. 2). Measured BTFs are assumed to be arranged like this in the following. For a very detailed overview of BTF acquisition devices, we refer the reader to the recent survey by Schwartz et al. [SSW*14].

3.2. Illumination multiplexing

Multiplexed measurement in general is based on the observation that by using series of appropriate patterns, the amount of *signal-independent* noise in the demultiplexed measurements will actually be lower than without multiplexing. In the special case of illumination multiplexing, the resulting linear system to be solved is

$$\mathbf{M} \cdot \mathbf{I}_{\text{single}} = \mathbf{I}_{\text{multiplexed}}$$

where $\mathbf{M} \in \mathbf{Z}^{n \times n}$, $\mathbf{I}_{\text{single}} \in \mathbf{R}^{n \times (w \cdot h)}$ with each row a $w \times h$ image of the scene lit by an individual light source, and $\mathbf{I}_{\text{multiplexed}} \in \mathbf{R}^{n \times (w \cdot h)}$ where each row is a $w \times h$ image of the scene lit by an individual illumination pattern determined by the corresponding row of $\mathbf{M}$.

In setups where light sources only have two states, "on" and "off", $\mathbf{M}$ needs to be binary. It has been shown that a class of binary matrices called *S-matrices* is at least very close to minimizing the average mean square error, and that the associated signal-to-signal-independent-noise ratio is increased by a factor of $\sqrt{n}/2$ for large n [HS79]. It has, however, also been shown that the presence of signal-dependent

© The Eurographics Association 2015.

noise such as photon noise strongly counteracts this advantage [HS79, SNB07].

A straight-forward way to obtain S-matrices $\mathbf{M} \in \mathbf{N}^{n \times n}$ of a given order $n = 4p+3$ for n, $p \in \mathbf{N}$ prime, is the *quadratic residue construction* [HS79]: Let Q be the set of *quadratic residues* of $0, \ldots, \frac{n-1}{2}$ modulo n. Then

$$m_{1j} = \begin{cases} 1, & j \in Q \\ 0, & \text{else} \end{cases}$$

The remaining $n-1$ rows of $\mathbf{M}$ are obtained by cyclically shifting the preceding row to the left by one.

If the number of light sources is not prime of the required form, which will likely be the case in setups not constructed with multiplexing in mind, one can choose a suitable higher order and truncate the resulting S-matrix such that the number of columns equals the number of light sources. The resulting linear system can then be solved in the least-squares sense [HS79].

4. Proposed method

4.1. Thresholded de-noising

As we wish to achieve the maximum gain with respect to compression of dynamic range and reduction of shutter times, we chose to use the highest S-matrix order applicable to our acquisition setup and, by extension, the largest amount of demultiplexing noise. In order to mitigate that noise, we assume the availability of a linear model $\mathbf{U}$ derived from a database $\mathbf{D}$ of traditionally measured BTFs by means of a truncated SVD

$$\mathbf{D} \approx \mathbf{U}\Sigma\mathbf{V}^t$$

of some rank $k \ll n$, which has been shown to be an optimal rank-k approximation in the least-squares sense [EY36]. Such models are known to generalize to materials $\mathbf{B}$ that do not belong to the particular database [dBSHK14]; i.e.

$$\min_{\mathbf{C_B}} ||\mathbf{UC_B} - \mathbf{B}|| < \varepsilon.$$

Den Brok et al. propose to de-noise by simply projecting a demultiplexed BTF, obtained as described in Sec. 3.2, onto the corresponding subspace via

$$\mathbf{B}_{\text{denoised}} = \mathbf{U} \cdot (\mathbf{U}^T \cdot \mathbf{B}_{\text{demultiplexed}})$$

We observed that demultiplexing produces a number of very noisy outlier textures that may have an undesirable impact on the projection. We therefore introduce one intermediate step to make this projection more robust: For the entire BTF, we compute per-texture variances and do not take the BTF's textures that correspond to some percentile of the variances into consideration during projection. Let $\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{B}}_{\text{demultiplexed}}$ be the linear model and demultiplexed BTF, respectively, with the rows corresponding to the selected textures removed. The equation thus becomes

$$\mathbf{B}_{\text{denoised}} = \mathbf{U} \cdot (\widetilde{\mathbf{U}}^{\dagger} \cdot \widetilde{\mathbf{B}}_{\text{demultiplexed}}),$$

where † denotes the Moore-Penrose pseudo-inverse. We tried various deciles with normalized and unnormalized data and found the 9th decile of the latter to perform best.

4.2. Database bootstrapping

A drawback of the the outlined method certainly is the assumed availability of a suitable database. We therefore propose a bootstrapping scheme that helps constructing a suitable database faster: Based on the observation that the amount of demultiplexing noise depends on both dynamic range and number of light sources, the idea is to divide the hemisphere of light sources into quadrants of approximately equal size, such that one quadrant contains the light sources directly opposite the cameras. As the direct reflections and Fresnel effect usually contribute most to the dynamic range, the images for this quadrant are best captured using single-light illumination. The images for the remaining three quadrants can be successively captured using appropriate S-matrix patterns. Unfortunately, this approach is somewhat tied to acquisition setup paradigms where such a division of the hemisphere is actually possible. Different means would need to be found for other paradigms.

5. Results

5.1. Experimental setup

We chose 12 materials from 3 classes – *cloth*, *leather*, and *wood* – each, of which we used the same four materials per class as den Brok et al. for performance evaluation. For the purpose of evaluating our database bootstrapping scheme we further selected four of the remaining leathers.

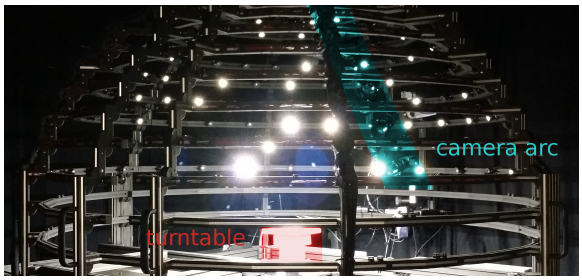


Figure 3: *Our measurement device illuminating a material sample with an S-matrix pattern.*

The measurement device we used in our experiments is a camera dome with 11 industrial-grade cameras with a maximum frame-rate of 8 Hz and 198 LED light sources, 10 of which are placed opposing the lower-most 10 cameras to produce direct reflections. The material sample is placed on a turntable that is rotated during acquisition to increase the setup's angular resolution. For an extensive description of the setup, cf. Schwartz et al. [SSWK13]. The quadratic residue construction requires the S-matrix order to be prime

© The Eurographics Association 2015.

of the form $4p+3$, p prime. The smallest S-matrix order of that form greater or equal the number of LEDs in our setup is $n = 199$, which we used in our experiments.

In order to speed up the acquisition of the ground truth data, the single-light measurements were obtained using a camera gain of 10 dB, while the multiplexed measurements were obtained with a camera gain of 0 dB. Additionally, we measured 4 small patches (approximately 1 cm $\times$ 1 cm) of different materials of the same class at once. We argue that, even in practive, this is a reasonable trade-off to be made: If required, larger BTFs for each single sample can be produced quickly by measuring them once more using the proposed method, or by extrapolation as described e.g. by Steinhausen et al. [SdBHK14].

The materials are placed on the turntable, which during measurement is rotated 12 times by 30°. For each turntable position, the cameras take pictures of the material lit by each of the LEDs separately using several exposure times determined manually before the actual acquisition. The test materials were additionally measured using S-matrix patterns of order $n = 199$ before proceeding to the next turntable position. In the case of leather, the four leathers selected for evaluation of the database bootstrapping scheme were additionally measured using S-matrix patterns of orders $n_{\{1,2,3\}} = 47$ for each of the three of the illumination hemisphere's quadrants not sampled using single-light illumination.

After measurement, we combine the low dynamic range (LDR) raw images into high dynamic range (HDR) images and subsequently demosaic and rectify them. We use LED calibration data in order to account for variations in the LEDs' spectra. For further details on the various post-processing steps, cf. Schwartz et al. [SSW*14].

Finally, the resulting images are cropped to minimize influence from the materials' surroundings and arranged and stored as matrices, which in the case of multiplexed measurements are then demultiplexed. For rendering we resample the BTFs in the angular domain such that the light and view hemispheres are the same for each texel.

5.2. De-noising

We compute the linear models intended for de-noising seperately from the per-color channel ABRDFs of each of the three material classes, leaving out one test material at a time. We use the log(Y) U/Y V/Y color space, which lends itself well to least-squares fitting [MPBM03, dBSHK14].

For our experiments we used 768 basis vectors for log(Y) and 128 for both U/Y and V/Y channels, which is sufficient for the material classes under consideration and allows for better comparability with the method proposed by den Brok et al. [dBSHK15].

Tab. 1 shows relative errors

$$\varepsilon = \frac{\|\mathbf{B}_{\text{reconstructed}} - \mathbf{B}_{\text{reference}}\|_{\text{F}}}{\|\mathbf{B}_{\text{reference}}\|_{\text{F}}}$$

for BTFs reconstructed in the denoted ways. Errors are computed on the log(Y) channel to approximate human perception. We found the general tendencies to be the same for the other color channels.

In the case of wood, the relative errors are relatively small, and indeed there are no obvious differences between the renderings of ground truth, demultiplexed and denoised BTFs, even though the relative error is reduced significantly by denoising (cf. Fig. 4, 1st row). Curiously, this is the only material class where the proposed method performs worse than the state of the art in purely numerical terms. We have yet to investigate what the reason might be. Possibly the error introduced by sparse reconstruction exceeds and hence does not counteract the low demultiplexing error. A higher threshold might mitigate this problem.

In contrast, Cloth 1 and Leather 2 exhibit annoying artifacts for grazing viewing angles, as seen at the cylinder's borders. The BTF de-noised using den Brok et al.'s method looks much more plausible than the demultiplexed BTF, but it still exhibits minor artefacts which are close to unnoticable in the BTF with our method (cf. Fig. 4, 2nd and 3rd row).

Demultiplexed Leather 4, however, exhibits so much noise (yielding a relative error of over 27%) that both denoising strategies break down (cf. Fig. 4, 4th row), even though the relative error is more than halved by den Brok et al's method and reduced even further by ours. Note that the intensity at grazing angles in the result produced using den Brok et al.'s method is much too low. Our method does not suffer from this problem; it fails, however, at reducing the demultiplexing noise to an acceptable level for this particular material.

As can be seen in Tab. 2, the larger amount of light reaching the material samples on each particular image makes it possible to reduce total acquisition times by about 75–95%, even though the ground truth data was obtained with a much higher camera gain. In the case of Leather 1–4 and Wood 1–4, our setup's minimum acquitision time is reached. For all materials, 4 different shutter times were necessary under single-light illumination, whereas with multiplexed illumination, only a single exposure step was needed in the case of leather and wood, and two in the case of cloth, the latter because the cloths exhibited significantly different albedos. As a result, storage requirements for the raw measured data are reduced by 50–75% from originally 0.5–1.5 TB.

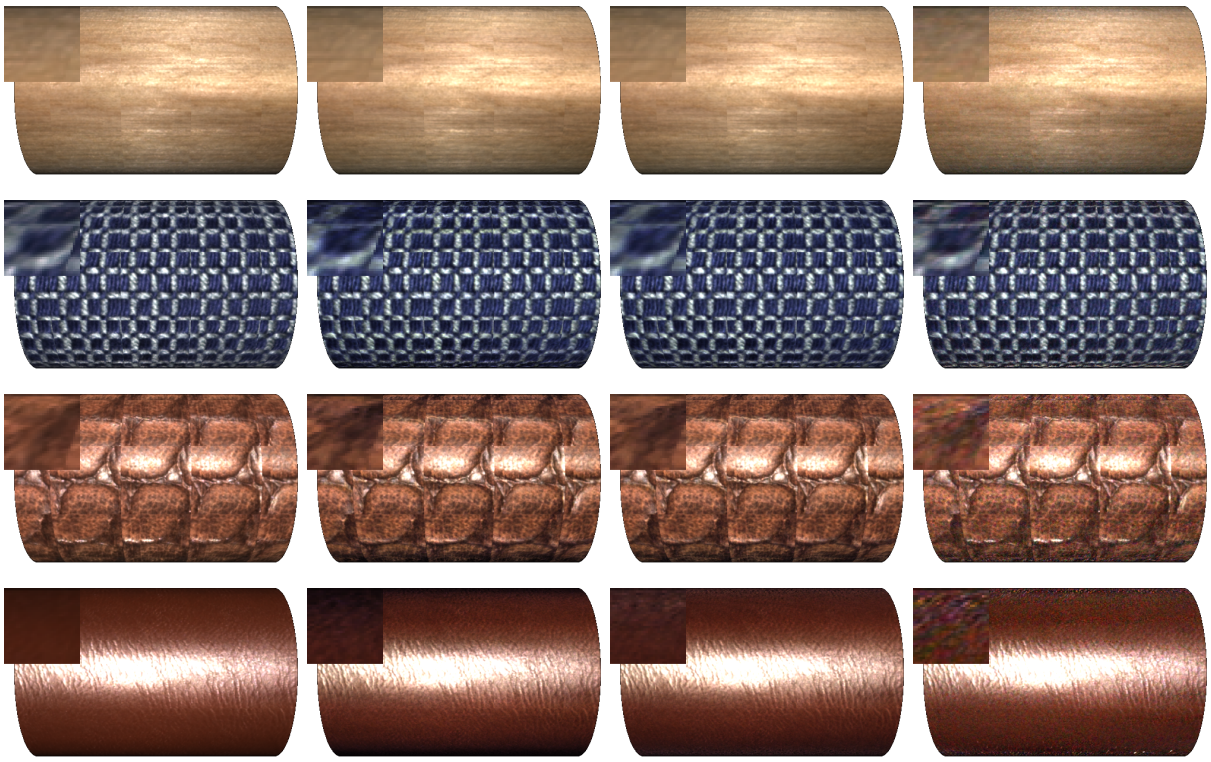
5.3. Database bootstrapping

We evaluated the performance of our database bootstrapping scheme in two steps. First, we compared the projection errors when projecting onto linear bases computed from the ground-truth and the bootstrapping measurements, respectively (cf. Tab. 3). We found projection errors to be only slightly bigger for the bootstrapping measurements. It is thus also to be expected that de-noising performance will

© The Eurographics Association 2015.

Class	$\varepsilon_{\text{demult}}$					$\varepsilon_{\text{denoised}}$					$\varepsilon_{\text{threshold}}$				
	1	2	3	4	∅	1	2	3	4	∅	1	2	3	4	∅
Cloth	16.2	12.2	10.5	11.3	12.6	7.5	6.2	5.8	5.6	6.3	5.3	4.5	4.3	4.7	4.7
Leather	17.7	20.1	23.0	27.4	22.1	8.8	8.3	9.4	10.5	9.3	6.3	7.5	7.8	7.6	7.3
Wood	7.3	7.6	9.3	7.2	7.9	3.6	3.7	4.2	3.7	3.8	4.9	5.1	4.2	4.3	4.6

Table 1: *Comparison of log(Y) relative errors [%], demultiplexed and denoised.*



(a) *Reference.* **(b)** *[dBSHK15]* **(c)** *Proposed method.* **(d)** *Demultiplexed.*

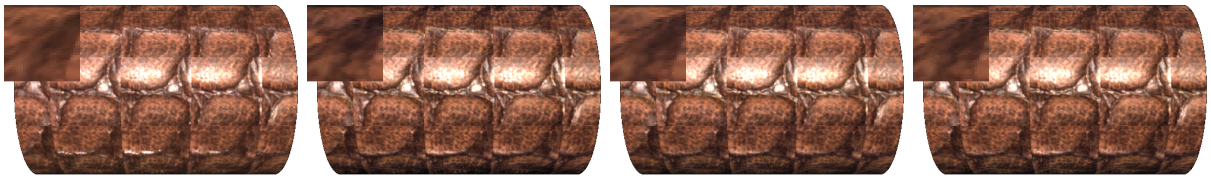
Figure 4: *Renderings of Wood 1, Cloth 1, Leather 2, and Leather 4.*

Class	shutter times [ms]		acquisition times [h]		rel. Δ
	single	multiplexed	single	multiplexed	
Cloth 1 – 4	150, 608.2, 2466.2, 10000	15, 45	10.8	1.8	−83%
Leather 1 – 4	10, 144.2, 2080.1, 30000	30	23.6	1.2	−95%
Wood 1 – 4	10, 60.4, 364.4, 2200	20	4.4	1.2	−75%

Table 2: *Comparison of shutter [ms] and acquisition times [h], single light vs. multiplexed.*

be slightly worse, which our experiments confirmed. However, both numerically and perceptually, we found it still to be better overall than that of den Brok et al.'s method, and the resulting BTFs are perceptually close to the those produced with the traditionally obtained basis (cf. Tab. 3 & Fig. 5). The acquisition itself took approximately 9 hours, as opposed to almost 24 hours for the ground-truth.

© The Eurographics Association 2015.



(a) *Reference.* **(b)** *[dBSHK15]* **(c)** *Bootstrapped basis.* **(d)** *Original basis.*

Figure 5: *Bootstrapping: renderings of Leather 4.*

	1	2	3	4	∅
$\varepsilon_{\text{proj}}$ ground-truth	8.4	7.6	8.6	9.3	8.5
$\varepsilon_{\text{proj}}$ bootstrapped	8.7	8.0	9.0	9.5	8.8
$\varepsilon_{\text{denoise}}$ ground-truth	6.3	7.5	7.8	7.3	7.3
$\varepsilon_{\text{denoise}}$ bootstrapped	7.0	8.3	8.8	8.5	8.2

Table 3: *Comparison of projection and reconstruction errors for Leather 1–4 with a bootstrapped database.*

6. Conclusion & Future work

We demonstrated the feasibility of illumination multiplexing in the context of BTFs, supported by using linear models derived from an existing database of material BTFs as a prior for a de-noising method that perceptually and, most of the time, numerically outperforms the state of the art. We found that using illumination multiplexing enables both dramatically reduced dynamic ranges and shutter times. As a result, storage requirements for raw measurement data could be reduced by up to 75%, and total acquisition times by up to 95%, even reaching the limits of our acquisition setup.

Moreover, we presented a "bootstrapping" method that allows for faster creation of a database suitable for the purpose of de-noising by using a hybrid approach where the quadrant of the illumination hemisphere likely to cause noise is sampled using single-light illumination, and the remaining quadrants are separately sampled using multiplexed illumination. That way, acquisition time could be reduced by about 63%. We found linear models derived from a database such obtained to perform not much worse than linear models derived from our ground-truth database.

It seems worthwhile to determine a proper noise model for our camera dome setup to use with the method proposed by Mitra et al. [MCV14] (cf. Sec. 2.1 for more details), which might help decrease the amount of demultiplexing noise even further.

Acknowledgements

This work was developed in the X-Rite Graduate School on Digital Material Appearance at the University of Bonn.

© The Eurographics Association 2015.

References

[dBSHK14] DEN BROK D., STEINHAUSEN H. C., HULLIN M. B., KLEIN R.: Patch-based sparse reconstruction of material BTFs. *Journal of WSCG 22*, 2 (June 2014), 83–90. 3, 4, 5

[dBSHK15] DEN BROK D., STEINHAUSEN C., HULLIN M. B., KLEIN R.: Multiplexed acquisition of bidirectional texture functions for materials. In *Measuring, Modeling, and Reproducing Material Appearance II (SPIE 9398)* (Feb. 2015), vol. 9398. 2, 5, 6, 7

[Don06] DONOHO D. L.: Compressed sensing. *IEEE Trans. Inform. Theory 52* (2006), 1289–1306. 2

[DvGNK99] DANA K. J., VAN GINNEKEN B., NAYAR S. K., KOENDERINK J. J.: Reflectance and texture of real-world surfaces. *ACM Trans. Graph. 18*, 1 (Jan. 1999), 1–34. 3

[DWT*10] DONG Y., WANG J., TONG X., SNYDER J., LAN Y., BEN-EZRA M., GUO B.: Manifold bootstrapping for SVBRDF capture. *ACM Trans. Graph. 29*, 4 (July 2010), 98:1–98:10. 3

[EY36] ECKART C., YOUNG G.: The approximation of one matrix by another of lower rank. *Psychometrika 1*, 3 (1936), 211–218. 4

[HS79] HARWIT M., SLOANE N.: *Hadamard transform optics*. Academic Press, 1979. 2, 3, 4

[KMBK03] KOUDELKA M. L., MAGDA S., BELHUMEUR P. N., KRIEGMAN D. J.: Acquisition, compression, and synthesis of bidirectional texture functions. In *ICCV Workshop on Texture Analysis and Synthesis* (2003). 2

[MCV14] MITRA K., COSSAIRT O., VEERARAGHAVAN A.: Can we beat hadamard multiplexing? data driven design and analysis for computational imaging systems. In *Computational Photography (ICCP), 2014 IEEE International Conference on* (May 2014), pp. 1–9. 2, 7

[MKU15] MIANDJI E., KRONANDER J., UNGER J.: Compressive image reconstruction in reduced union of subspaces. In *Eurographics 2015* (May 2015). 3

[MPBM03] MATUSIK W., PFISTER H., BRAND M., MCMILLAN L.: A data-driven reflectance model. *ACM Trans. Graph. 22*, 3 (July 2003), 759–769. 2, 3, 5

[MWBR13] MARWAH K., WETZSTEIN G., BANDO Y., RASKAR R.: Compressive light field photography using overcomplete dictionaries and optimized projections. *ACM Trans. Graph. 32*, 4 (July 2013), 46:1–46:12. 3

[PML*09] PEERS P., MAHAJAN D. K., LAMOND B., GHOSH A., MATUSIK W., RAMAMOORTHI R., DEBEVEC P.: Compressive light transport sensing. *ACM Trans. Graph. 28*, 1 (Feb. 2009), 3:1–3:18. 2

[RS07] RATNER N., SCHECHNER Y.: Illumination multiplexing within fundamental limits. In *Computer Vision and Pattern*

Recognition, 2007. CVPR '07. IEEE Conference on (June 2007), pp. 1–8. 2

[SdBHK14] STEINHAUSEN H. C., DEN BROK D., HULLIN M. B., KLEIN R.: Acquiring bidirectional texture functions for large-scale material samples. *Journal of WSCG 22*, 2 (June 2014), 73–82. 5

[SNB07] SCHECHNER Y. Y., NAYAR S. K., BELHUMEUR P. N.: Multiplexing for optimal lighting. *IEEE Trans. Pattern Anal. Mach. Intell. 29*, 8 (Aug. 2007), 1339–1354. 2, 4

[SSW*14] SCHWARTZ C., SARLETTE R., WEINMANN M., RUMP M., KLEIN R.: Design and implementation of practical bidirectional texture function measurement devices focusing on the developments at the university of bonn. *Sensors 14*, 5 (Apr. 2014). 3, 5

[SSWK13] SCHWARTZ C., SARLETTE R., WEINMANN M., KLEIN R.: DOME II: A parallelized BTF acquisition system. In *Eurographics Workshop on Material Appearance Modeling: Issues and Acquisition* (June 2013), Eurographics Association, pp. 25–31. 4

[WGT*05] WENGER A., GARDNER A., TCHOU C., UNGER J., HAWKINS T., DEBEVEC P.: Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Trans. Graph. 24*, 3 (July 2005), 756–764. 2

[WHON97] WONG T.-T., HENG P.-A., OR S.-H., NG W.-Y.: Image-based rendering with controllable illumination. In *Proceedings of the Eurographics Workshop on Rendering Techniques '97* (London, UK, UK, 1997), Springer-Verlag, pp. 13–22. 3

© The Eurographics Association 2015.

APPENDIX D

Rapid material capture through sparse and multiplexed measurements

Computers & Graphics 73 (2018) 26–36

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Technical Section

Rapid material capture through sparse and multiplexed measurements☆

Dennis den Brok*, Michael Weinmann, Reinhard Klein

Institute of Computer Science II, University of Bonn, Endenicher Allee 19a, Bonn 53115, Germany

ARTICLE INFO

Article history:
Received 11 October 2017
Revised 9 February 2018
Accepted 6 March 2018
Available online 20 March 2018

Keywords:
Appearance acquisition
Sparse acquisition
Illumination multiplexing
Appearance modeling

ABSTRACT

Among the many models for material appearance, data-driven representations like bidirectional texture functions (BTFs) play an important role as they provide accurate real-time reproduction of complex light transport effects such as interreflections. However, their acquisition involves time-consuming capturing of many thousands of bidirectional samples in order to avoid interpolation artifacts. Furthermore, high dynamic range imaging including many and long exposure steps is necessary in the presence of low albedo or self-shadowing. So far, these problems have been dealt with separately by means of sparse reconstruction and multiplexed illumination techniques, respectively. Existing methods rely on data-driven models learned on data that has been range-reduced in a way that made their simultaneous application impossible. In this paper, we address both problems at once through a novel method for learning data-driven appearance models, based on moving the dynamic range reduction from the data to the metric. Specifically, we learn models by minimizing the relative L_2 error on the original data instead of the absolute L_2 error on range-reduced data. We demonstrate that the models thus obtained allow for faithful reconstruction of material appearance from sparse and illumination-multiplexed measurements, greatly reducing both the number of images and the shutter times required. As a result, we are able to reduce acquisition times down to the order of minutes from what used to be the order of hours.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Analytical material reflectance models such as (spatially varying) bidirectional reflectance distribution functions ((SV)BRDFs) [1], which model reflectance per surface point depending on the incident and outgoing light directions, can nowadays be obtained efficiently, as simply as by taking two photographs with a cellphone camera [2]. Many applications require a higher degree of accuracy than what these models are able to deliver, or real-time rendering including meso-scale light transport effects that require solving global illumination, e.g. interreflections and self-shadowing. Image-based representations such as data-driven SVBRDFs or bidirectional texture functions (BTFs) [3] provide these advantages. Capturing them, however, demands much more effort, up to days for a single material [4]. This can largely be attributed to two factors: high-frequency features and dynamic range. The former can be caused e.g. by specularity, parallax and shadows. In order to avoid visible sub-sampling artifacts in rendering, often tens of thousands of images need to be obtained. The latter is a consequence particularly of specularity, but also of shadows or low albedo. The more prominent these effects, the greater the number of exposure steps and the maximum exposure time necessary to capture a material's full dynamic range.

By now there are a number of approaches to solving these problems separately. Sparse acquisition techniques are applied when only a small subset of the desired dense sampling is actually measured; the remaining data is then interpolated by means of linear models learned from an existing material database (e.g. den Brok et al. [5], Nielsen et al. [6]). Conversely, illumination multiplexing exploits the linearity of the superposition of light by illuminating the material sample not with a single light source but with patterns of light sources, which increases the amount of light on the sample and eliminates shadows, thereby reducing dynamic range (see Fig. 3). The desired representation with one active light source per image can then be obtained by solving an appropriate linear system, a process that is, however, known to be detrimental to the signal-to-noise ratio (SNR). The models used in sparse acquisition have been shown to also help mitigate the noise problems [7].

Either way, acquisition times can be reduced significantly, down to the range of at most a few hours [4], but still far from what acquisition devices for analytical SVBRDFs are capable of. As the approaches are completely orthogonal, the question arises whether

☆ This article was recommended for publication by H. Rushmeier.

* Corresponding author.
E-mail address: denbrok@cs.uni-bonn.de (D. den Brok).

https://doi.org/10.1016/j.cag.2018.03.003
0097-8493/© 2018 Elsevier Ltd. All rights reserved.

the two paradigms can be combined. So far, this has been impossible: the linear bases used as models in the above approaches rely heavily on range-reduction techniques such as logarithmic transformations applied to the training data. These transformations, however, do not commutate with multiplexing; i.e., we cannot infer the transformed data from a multiplexed measurement without prior de-multiplexing. But de-multiplexing requires images for *all* multiplexing patterns, which we wish to avoid in sparse acquisition.

In this paper, we present, to the best of our knowledge, the first approach to accurate reflectance acquisition which simultaneously exploits sparse acquisition and multiplexed illumination, enabling faithful BTF acquisition in several minutes. Specifically, we propose a different approach to dynamic range reduction in model learning: rather than the *absolute* L_2 error on non-linearly transformed data as a metric, we minimize the *relative* L_2 error on untransformed data, which ultimately allows for sparse multiplexed acquisition of BTFs. As obtaining a basis this way is not as straightforward as simply computing a truncated singular value decomposition (SVD), we provide an efficient alternating least-squares approach to compute a suitable basis. As demonstrated by our results, combined sparse and multiplexed acquisition allows for a reduction of acquisition time from the order of hours/days required for brute-force measurements down to only several minutes, significantly outperforming both sparse acquisition and multiplexed acquisition.

We evaluate the performance in the sparse and multiplexed case, both separately and combined, and compare against the state-of-the-art. In our evaluation, we find that a method recently presented by Nielsen et al. [6], which had only been tested on BRDFs and flat SVBRDFs so far, also works on material BTFs and slightly outperforms the state-of-the-art in this field.

In summary, our paper presents the following key contributions:

- a novel basis for measured material appearance based on minimizing the *relative* L_2 error on the *untransformed* data instead of the *absolute* L_2 error on non-linearly transformed data,
- an evaluation of our basis' performance as a model for the appearance of typical real-world materials in the context of sparse or multiplexed acquisition.
- *En passant*, we find that a recently presented sparse acquisition method only known so far to work for BRDFs and flat SVBRDFs also lends itself to arbitrary material BTFs and slightly outperforms the state-of-the-art in this field.
- We demonstrate that our basis is designed to take advantage of both sparse acquisition and multiplexed illumination at once, resulting in an overall acquisition speed-up of up several orders of magnitude in comparison to a full measurement, and a still significant speed-up of the acquisition process in comparison to sparse or multiplexed acquisition, while maintaining perceptually accurate results.

2. Related work

In this section, we briefly review related work on modeling surface appearance including fine surface details. Furthermore, we discuss previous work on fast appearance acquisition based on the aforementioned concepts of sparse acquisition and illumination multiplexing.

2.1. Acquisition and modeling of material appearance

Detailed surveys on appearance acquisition and modeling can be found in the literature [8–11]. Widely used reflectance models that capture spatially varying material characteristics under varying viewing and illumination conditions include spatially varying bidirectional reflectance distribution functions (SVBRDFs) [1] and bidirectional texture functions (BTFs) [3]. In contrast to SVBRDFs, BTFs are not necessarily defined with respect to the material's actual surface. Indeed, often a planar reference geometry is assumed, as for many materials like irregular fabrics it is difficult or practically impossible to determine the exact surface geometry with conventional acquisition setups. As a result, SVBRDFs do not accurately capture the light exchange for such materials. Moreover, BTFs do not impose restrictions regarding energy conservation on the per-texel BRDFs and simply encode the *appearance* of the material at one particular coordinate on the reference geometry, which is why they are known as *apparent* BRDFs (ABRDFs) [12]. Together with the parametrization over a flat geometry this allows capturing non-local effects such as interreflections, self-occlusions and self-shadowing. As measured SVBDRFs can be considered a subclass of BTFs, we shall focus on BTFs in this work. Due to their generality, BTFs are impossible to model, which is why one typically retreats to image-based representations that can be evaluated through a (possibly interpolated) table look-up. Measured BTFs have natural representations as matrices $\mathbf{B} \in \mathbf{R}^{n_{lv} \times n_{tx}}$, where the rows correspond to linearized light- and view-dependent 2D textures, the columns to linearized per-texel apparent BRDFs (cf. Fig. 2), n_{tx} denotes the number of texels (incorporating color channels for brevity) and n_{lv} the number of pairs of incoming and outgoing light directions under consideration. Note that in order to avoid interpolation artifacts, it is desirable that n_{lv} be large, in the order of thousands or tens of thousands, which in practice translates to the expensive process of acquiring tens of thousands of images of a given material. Given the matrix representation, both existing methods to mitigate this problem and the proposed method can be written concisely in terms of matrix operations, as we shall detail on in the following.

2.2. Sparse reflectance acquisition

Seminal work on sparse reflectance acquisition has been introduced by Matusik et al. [13] with the introduction of a new reflectance model that represents materials in terms of linear combinations from a set of densely sampled BRDF measurements. In subsequent work, Matusik et al. [14] approach the sparse reconstruction of isotropic BRDFs based on both a wavelet basis and a linear model obtained from the MERL database of isotropic BRDFs. They, however, did not investigate generalizations of these results to more complex reflectance models. In a closely related work, BTF compression was approached by Koudelka et al. [15], where single linear models for apparent BRDFs have been computed per-material. So far, the only technique that focuses on sparse reconstruction of BTFs has been presented by den Brok et al. [5]. Similar to the technique presented by Matusik et al. [13], a linear model is derived by applying singular value decomposition on a (logarithmically transformed) database of ABRDFs. By fitting these models to small BTF patches, non-local effects of light exchange are taken into account and BTFs have been reconstructed from only 6% of the typically used view-light conditions without a reduction of the resolution determined by the acquisition setup. Furthermore, manifold bootstrapping has been introduced by Dong et al. [16], where a manifold is constructed from analytical BRDFs fitted to BRDF measurements of certain manually selected surface positions on the material sample and used for the reconstruction of anisotropic SVBRDFs from sparse measurements. While a generalization to previously unseen materials might be achieved, the significant increase of the dimensionality of the manifold of per-texel reflectance distribution functions makes this technique impractical for BTFs. Nielsen et al. [6] present an approach for BRDF reconstruction from an optimized sparse sampling, where an improved logarithmic mapping of the MERL database is employed to obtain

a simple linear BRDF subspace. This technique has been extended by Xu et al. [17], where an improved minimal sampling method for data-driven reflectance acquisition of homogeneous, flat materials from two images using the MERL BRDF database as prior has been presented. One further extension has been proposed by Yu et al. [18], who independently apply the technique of Nielsen et al. to each pixel/point on the material. The technique presented by Zhou et al. [19] exploits the sparse basis in terms of a convex combination over a limited number of target-specific basis materials and sparse blend priors for SVBRDF acquisition from a small number of images or even a single image. In particular, the relative reconstruction error of the measurements is minimized while encouraging a small number of representatives and a blending between a small number of them. A constrained optimization is applied where the representatives are represented as a linear combination of generic BRDFs which is solved iteratively based on an alternating application of quadratic programming to the representatives and the blending weights.

Further investigations focused on the use of compressed sensing techniques. Peers et al. [20] apply compressed sensing for the acquisition of reflectance fields based on a hierarchical, multiresolution Haar wavelet basis that exploits spatial coherence. The assumption of incident 2D light fields with fixed view conditions and outgoing 2D light fields, however, cannot directly be extended to multi-view scenarios as given for BTFs. Furthermore, the limited number of light source positions provided by parallel BTF acquisition devices counteracts the advantage of compressed sensing and shot noise might be introduced.

Focusing on the sparse reconstruction of light fields, the techniques presented by Marwah et al. [21] and Miandji et al. [22] rely on dictionary based sparse representation of the measured light field and non-linear sparse coding techniques are applied. As pointed out by den Brok et al. [5], such a dictionary is likely to not exist for ABRDFs due to their high dimensionality.

Further compressive sensing techniques focus on hand-held light field acquisition with commodity hardware ([23,24]).

Aittala et al. [25] propose (analytical) SVBRDF acquisition for flat materials based on a setup that involves a single LCD screen and a camera. Continuous illumination patterns and a specially designed image formation model allows SVBRDF acquisition from relatively few measurements. The trend of hand-held acquisition recently culminated in the extremely light-weight approaches presented by Aittala et al. [2,26].

Aittala et al. [2] focus on the acquisition of coarse anisotropic SVBRDFs for flat, stationary materials that exhibit self-similarities, i.e. there are several points on the texture with similar reflectance properties. This allows for gathering different reflectance samples across the materials based on taking two approximately fronto-parallel images taken under ambient illumination and flash illumination and the subsequent fitting of an analytical SVBRDF model.

Based on similar simplifying assumptions, the follow-up work [26] estimates statistically similar material appearance in terms of SVBRDFs from a single head-lit flash image based on neural texture synthesis. While plausible material reconstructions are shown, these do not represent accurate reconstructions of the real-world materials.

2.3. Multiplexed acquisition

In their theoretical work, Harwit and Sloane [27] proof the optimality of Hadamard patterns in terms of reduction of signal-independent noise. This is accompanied by a discussion of sources of noise that occur in optical systems as well as their influence on demultiplexing. Based on this work, Wenger et al. [28] used multiplexing with different pattern types (triangles of lights and Hadamard patterns) for the acquisition of time-varying light fields of human faces. In particular, Hadamard patterns have been demonstrated to result in noisy results that cannot be improved by straight-forward filtering. Schechner et al. [29] also propose a multiplexing approach based on Hadamard patterns and additionally provide an analysis of different noise types introduced by stray light, saturation, and noisy illumination sources. Furthermore, a criterion is provided that indicates whether illumination multiplexing is beneficial for a certain setup. To deal with photon noise and sensor saturation which impact Hadamard-based multiplexing schemes, Ratner and Schechner [30] derive multiplexing codes that are still optimal under these effects. However, the assumptions of a 1D affine noise model and a nearly diffuse scene are not valid for materials exhibiting more complex effects of light exchange.

Park et al. [31] focus on continuous spectral reflectance reconstruction for each scene point from spectral measurements and apply illumination multiplexing based on a linear model to reduce the number of required measurements. Liang et al. [32] address the limitations of Hadamard code-based multiplexing in the case of shot noise with an optimized multiplexing scheme for programmable aperture photography of light fields. More recently, Mitra et al. [33] exploit image priors for the optimization of the illumination patterns. This allows for dealing with large amounts of light. However, the assumption of a 1D affine noise model as well as the computational burden of the extension to low-resolution light fields make this approach impractical for BTF acquisition. den Brok et al. [7] presented the first approach of illumination multiplexing for BTF acquisition. By exploiting the linearity of light transport, images for the individual light-view configurations are obtained from the images acquired under illumination multiplexed with Hadamard patterns by solving an appropriate linear system. Noise contained in the demultiplexing result depending on the dynamic range of the considered material is removed by projecting the noisy BTF onto a linear subspace obtained from a database of conventionally measured BTFs.

Miyagawa and Taniguchi [34] focus on the recovery of dense and accurate light transports from objects. For this purpose, orthogonal illumination based on a Walsh–Hadamard matrix is used for relighting which allows to consider ambient illumination in addition to directly reflected light.

3. Preliminaries

In this section, we discuss basics on sparse acquisition and multiplexed acquisition as prerequisites of the following sections:

3.1. Sparse acquisition

In the context of material acquisition, sparse acquisition typically refers to taking only a small subset of the total amount of images. This can be formulated in terms of considering a matrix product $\mathbf{SB}$ instead of the full measurement $\mathbf{B}$, where the binary sparse measurement matrix $\mathbf{S} \in \{0, 1\}^{n_s \times n_{lv}}$, $\mathbf{S}\mathbf{S}^T = \mathbf{1}$, selects the desired rows of $\mathbf{B}$. Under the assumption that $\mathbf{B}$ can be approximated well by a linear basis $\mathbf{U}$, a standard way to obtain an approximation $\mathbf{B} \approx \mathbf{U}\mathbf{V}$ is

$$\mathbf{V} = \operatorname{argmin}_{\tilde{\mathbf{V}}} \|\mathbf{S}\mathbf{U}\tilde{\mathbf{V}} - \mathbf{S}\mathbf{B}\|_F^2 + \|\mathbf{R}\tilde{\mathbf{V}}\|_F^2, \tag{1}$$

where the second summand is an optional Tihonov regularization term supposed to penalize implausible solutions. In the case of BRDFs, Nielsen et al. [6], for instance, propose to take $\mathbf{R} = \Sigma^{-1}$, where Σ is the diagonal matrix of singular values corresponding to the singular vectors in $\mathbf{U}$, which penalizes large deviations from the training set's distribution of basis coefficients. This had so far not been investigated on BTF data, but in the matrix representation given in Section 2.1, it readily applies to BTFs as well. In contrast, den Brok et al. [5] actually reconstruct BTFs, but using a slightly

different matrix representation consisting of small BTF patches, i.e. several neighboring apparent BRDFs stacked on top of each other, which acts as regularization. As we find the method proposed by Nielsen et al. to perform better, we leave out further details.

3.2. Multiplexed illumination

A well-established approach to reduce acquisition times in setups with multiple light sources is provided by illumination multiplexing, i.e. by the simultaneous illumination of a material sample by several light sources following certain illumination patterns. Thereby, the amount of light illuminating the scene is increased which allows a reduction of the shutter times required to obtain the images (cf. Fig. 3). Moreover, in typical BTF acquisition setups, light sources are distributed on almost the entire hemisphere above the material sample. Regions that would be shadowed under illumination by a single light source will thus typically be lit in the multiplexing case, allowing for a further decrease of shutter times.

Formally, multiplexing corresponds to measuring **MB**, where $\mathbf{M} \in \{0, 1\}^{n_p \times n_{lv}}$ denotes a multiplexing matrix that specifies the n_p illumination patterns used during acquisition. By choosing **M** as an invertible matrix, **B** can be reconstructed from **MB**. However, for imaging systems like BTF acquisition setups, measurements are slightly distorted by Poisson noise, and, as a result, the reconstruction process is known to produce possibly intolerably noisy reconstructions [27]. den Brok et al. [7] demonstrated that in the case of material BTFs Hadamard illumination patterns provide a good trade-off between the number of simultaneous light sources and noise level in that the otherwise intolerable noise can be dealt with by the projection to a basis **U** computed on logarithmically transformed data:

$$\mathbf{V} = \operatorname{argmin}_{\tilde{\mathbf{V}}} \|\mathbf{U}\tilde{\mathbf{V}} - \mathbf{M}^{\dagger}(\mathbf{MB})\|_{\mathrm{F}}^2, \tag{2}$$

where $\mathbf{M}^{\dagger}$ denotes the Moore-Penrose pseudo-inverse of **M**, often resulting in perceptually good approximations $\mathbf{B} \approx \mathbf{UV}$.

3.3. Basis acquisition and computation

So far, we have not specified what the bases **U** in the previous sections look like and how to obtain them. Given a database **D** of measured BTFs, where the separate BTFs are concatenated along the second dimension, i.e. D is a matrix with ABRDFs as columns, a straight-forward approach is the use of matrix factorization techniques like the (truncated) singular value decomposition (SVD), e.g.

$$\mathbf{D} \approx \mathbf{U}\Sigma\mathbf{V}^{T}.$$

Matusik et al. [13] have shown that the dynamic range of the reflectance data, determined by the huge difference in brightness between specular highlights and diffuse reflection which may amount to several orders of magnitude, leads to a severe overfitting of the highlights. A number of metrics have been proposed to overcome this problem. In their recent work, Nielsen et al. [6] propose to use

$$\varrho \mapsto \log\left(\frac{\varrho \cos_{\text{weight}} + \varepsilon}{\varrho_{\text{ref}} \cos_{\text{weight}} + \varepsilon}\right), \tag{3}$$

where ϱ denotes a single (A)BRDF, ϱ_{ref} a reference (A)BRDF, taken as the median (A)BRDF,

$$\cos_{\text{weight}} := \max\{\cos\vartheta_i \cos\vartheta_o, \varepsilon\}$$

and $\varepsilon = 0.0001$ is a small offset to avoid division by zero. Note that Nielsen et al. [6] and derivative works like Vavra and Fili [35] and Yu et al. [18] validate this method on isotropic, anisotropic and flat (SV)BRDFs only, respectively.

The MERL BRDF database used in most prior works only contains 100 isotropic BRDFs which can reasonably be assumed to be representative of at least a large subspace of homogeneous isotropic materials. In contrast, the space of arbitrary ABRDFs is much larger, and many sample ABRDFs are required in order to obtain a representative basis. A CPU- and memory-efficient way is to compute bases on a per-material basis and merging them using methods like eigenspace merging as proposed by den Brok et al. [5].

4. Our approach

While the aforementioned techniques regarding sparse acquisition and illumination multiplexing individually allow a significant speed-up of the acquisition process by several orders, the acquisition times are still rather long in comparison to conventional SVBRDF acquisition. By considering the success of applying the individual techniques, the question arises whether the combination of these approaches is possible and allows an even more significant speed-up of the acquisition resulting in practical acquisition times of several minutes instead of hours. For this purpose, we propose to pose the combination of the different paradigms of sparse acquisition and multiplexed acquisition as an optimization problem

$$\mathbf{V} = \operatorname{argmin}_{\tilde{\mathbf{V}}} \|\mathbf{SMU}\tilde{\mathbf{V}} - \mathbf{SMB}\|_{\mathrm{F}}^2 + \|\mathbf{R}\tilde{\mathbf{V}}\|_{\mathrm{F}}^2, \tag{4}$$

where the first summand represents the data term and the second summand an optional regularization term as described in Section 3.1. This corresponds to obtaining a possibly small number of images of the material sample lit by and viewed from selected illumination patterns and camera positions, respectively. It turns out that the obstacle is the very idea which makes the previous approaches practical in their domains: the application of a logarithmic scaling to the data. The basis **U** typically is a basis for log-space data, but we cannot infer $\mathbf{M}\log(\mathbf{B})$ from the measurement **MB**, because in general $\log(\mathbf{MB}) \neq \mathbf{M}\log(\mathbf{B})$. We thus propose to use a different strategy to deal with the data's dynamic range. The respective details are discussed in the following:

4.1. Relative error metric

Inspired by Ruiters et al. [36], we modify the metric used when computing the basis by assigning per-entry weights **W** to the L_2 errors instead of modifying the data. The optimization problem to be solved then becomes

$$\mathbf{U}, \mathbf{V} = \operatorname{argmin}_{\tilde{\mathbf{U}}, \tilde{\mathbf{V}}} \|\mathbf{W} \odot (\tilde{\mathbf{U}}\tilde{\mathbf{V}} - \mathbf{D})\|_{\mathrm{F}}, \tag{5}$$

where $\odot$ denotes the entry-wise matrix product. By taking **W** as the entry-wise inverse of **D**, this is equivalent to minimizing the *relative* L_2 error instead of the absolute one ([36]). Note that it is still possible to modify **W** to include additional weights such as the cosine weighting mentioned above, or the additional sub-weights proposed by Bagher et al. [37], by simply multiplying them with the proposed weights, or replacing them altogether.

4.2. Basis computation

As in related work, we assume the availability of a database $\{\mathbf{D}^{(i)} \in \mathbf{R}^{n_{lv} \times n_{tx}} \mid 1 \leq i \leq n\}$ of n fully measured material BTFs. Like den Brok et al. [5], we compute bases per-material first and later merge the results into a single basis. As, to the best of our knowledge, there is no canonical way to do the former, we chose to apply an alternating least-squares approach originally inspired by tensor rank decomposition which allows for per-entry weights [38].

For readability, denote by **U**, **V** the two factors approximating $\mathbf{D}^{(i)}$. Let c be the approximate rank of $\mathbf{D}^{(i)}$ and $\mathbf{d}^{(i)}$, $\mathbf{w}$ be the vectors

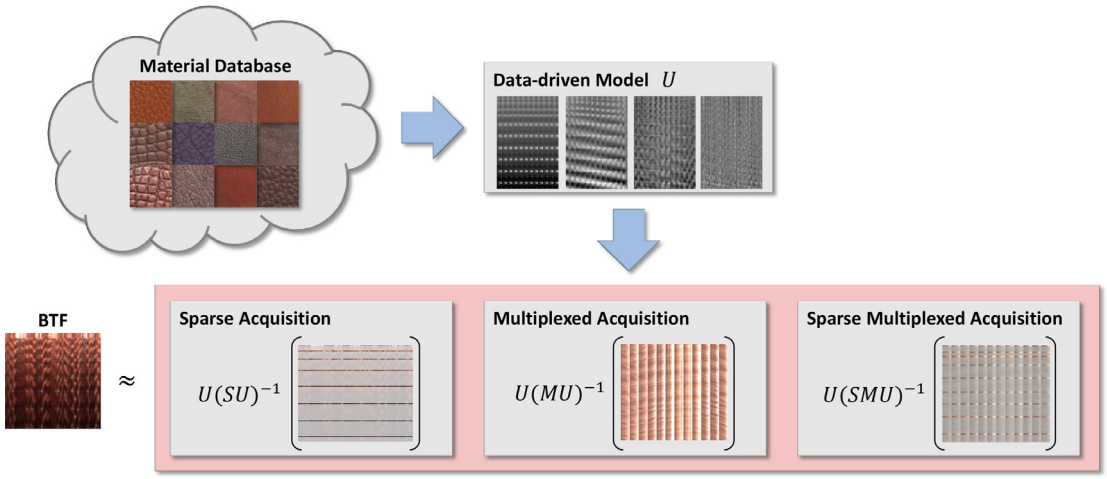


Fig. 1. Sketch of the proposed method: given a database of traditionally measured BTFs, a data-driven linear model **U** is derived in a novel way that allows for both sparse reconstruction and illumination-multiplexed acquisition to be used separately as well as, for the first time, simultaneously.

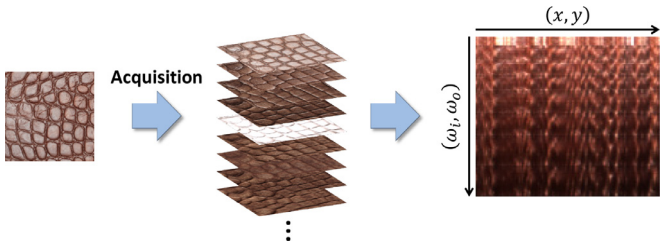


Fig. 2. Representation of a discretized BTF as a matrix. (x, y) denotes the spatial coordinates in the BTF patch, i.e. there are n_{tx} columns in the matrix. The number of rows n_{lv} corresponds to the number of view-light configurations (ω_i, ω_o).

of entries of $\mathbf{D}^{(i)}$ and $\mathbf{W}$, respectively. We initialize $\mathbf{U} \in \mathbf{R}^{n_{lv} \times c}$ with random values drawn uniformly from the interval $(-\frac{1}{2}, \frac{1}{2})$. Then, we may determine $\mathbf{V} \in \mathbf{R}^{c \times n_{tx}}$ by solving for its vector $\mathbf{v}$ of entries as

$$\mathbf{v} = \operatorname{argmin}_{\tilde{\mathbf{v}}} \| \operatorname{diag}(\mathbf{w}) \operatorname{diag}(\mathbf{U}, \ldots, \mathbf{U})\, \tilde{\mathbf{v}} - \operatorname{diag}(\mathbf{w})\, \mathbf{d}^{(i)} \|_2, \tag{6}$$

where $\operatorname{diag}(\mathbf{U}, \ldots, \mathbf{U})$ is a block-diagonal matrix of n_{tx} copies of $\mathbf{U}$. Although the involved matrices are very sparse, the resulting linear system is still impractical to handle. Fortunately, the matrices' sparsity structure allows for the system to be solved per-texel, possibly in parallel, via

$$\mathbf{V}_{:,x} = \left((\mathbf{U}^T \operatorname{diag}(\mathbf{W}_{:,x})^2\, \mathbf{U} \right)^{-1} \mathbf{U}^T \operatorname{diag}(\mathbf{W}_{:,x})^2\, \mathbf{D}^{(i)}_{:,x}. \tag{7}$$

Similarly, given a new estimate of **V**, **U** can be obtained via

$$\mathbf{U}_{lv,:} = \left((\mathbf{V} \operatorname{diag}(\mathbf{W}_{lv,:})^2\, \mathbf{V}^T \right)^{-1} \mathbf{V} \operatorname{diag}(\mathbf{W}_{lv,:})^2\, \mathbf{D}^{(i)}_{lv,:}. \tag{8}$$

These alterning optimizations are iterated until convergence, and finally $\mathbf{U}^{(i)}$ is taken as the normalized columns of **U**, and $\Sigma^{(i)}$ as the diagonal matrix of products of column and row norms of **U** and **V**, respectively. Finally, a basis **U** for the entire database $\{\mathbf{D}^{(i)}\}$ can be obtained by computing the SVD

$$\mathbf{U} \Sigma \mathbf{V}^T = \left(\mathbf{U}^{(1)} \Sigma^{(1)}, \ldots, \mathbf{U}^{(n)} \Sigma^{(n)} \right), \tag{9}$$

which is cheap when the database is relatively small, or by using techniques like eigenspace merging [39]. Besides **U**, we also store Σ for the purpose of regularization. Cf. Fig. 1 for a brief overview. Note that this algorithm is not guaranteed or even likely to converge against the global minimum. The minima it found in our experiments, however, always worked reasonably well.

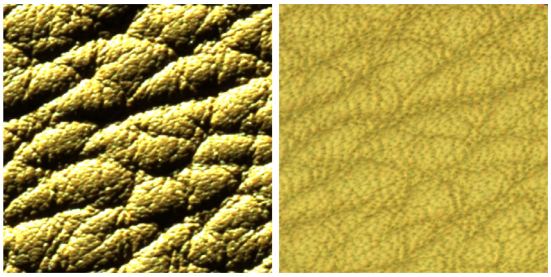


Fig. 3. Material sample lit by single light source (left; scaled by factor 50) and Hadamard illumination pattern (right; no scaling applied). Note the right image's flattened dynamic range and greatly increased brightness, which drastically reduces shutter times and exposure steps required.

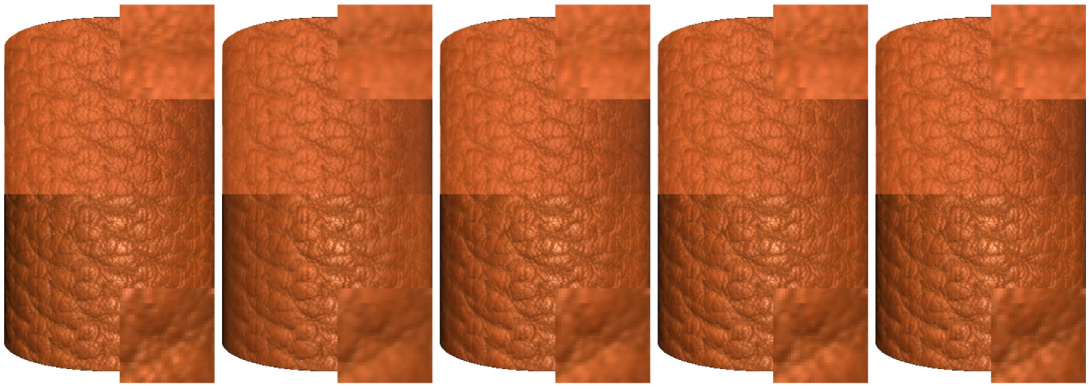
4.3. Sampling strategy

Nielsen et al. [6] propose a sophisticated optimization algorithm to determine a specified number n of coordinates to sample, assuming the samples lie on a regular grid defined on the 3D Rusinkiewicz parameterization [40]. As this assumption is violated by typical BTF acquisition setups, we use the more general but less effective algorithm originally proposed by Matusik et al. [13]. The goal is to find a measurement matrix $\mathbf{S} \in \{0, 1\}^{n_s \times n_{lv}}$ with $\mathbf{S}\mathbf{S}^T = \mathbf{1}$ such that the condition number $\kappa(\mathbf{SMU}\Sigma)$ becomes minimal, because this is a good indicator that redundant sample coordinates have been avoided. Here, n_s denotes the number of sparse samples. A greedy strategy is to start with a random subsampling and iteratively testing whether exchanging a random coordinate with another leads to a smaller condition number, and, if so, keeping the new coordinate, until convergence or until a time-limit is reached.

Note that once **M** encodes illumination patterns with many simultaneous light sources, the dynamic range of **MU** may be greatly reduced, and consequently the condition numbers do not vary as much with the chosen subsampling. We expect this to be an indicator that in the case of multiplexed illumination the choice of subsampling is less important. We did not investigate whether the choice of patterns is irrelevant altogether, but the sampling strategy needs to be determined once per basis and desired sparsity only, anyway.

4.4. Reconstruction

Once a sparse, multiplexed measurement **SMB** has been obtained, determining an approximation $\mathbf{B} \approx \mathbf{UV}$ is straight-forward:



(a) Reference. (b) $d = 256$, PSNR 33.3 dB (c) $d = 512$, PSNR 36.2 dB (d) $d = 768$, PSNR 38.5 dB (e) $d = 1024$, PSNR 40.5 dB

Fig. 4. Basis projections for various numbers d of basis vectors, ours. Top half: environmental lighting, bottom half: point light. Typically, $d = 1024$ yields good results.

$$\mathbf{V} = \left((\mathbf{SMU})^T(\mathbf{SMU}) + \mathbf{R}^T\mathbf{R})\right)^{-1}(\mathbf{SMU})^T(\mathbf{SMB}), \tag{10}$$

where $\mathbf{R}$ is an appropriate regularization matrix. One might assume that this projection would also benefit from weighting, which could be achieved using an alternating least-squares approach as above. We did not find any significant differences between the algorithms' results in practice and thus opted for the much faster one (Eq. (10)). Similar to Nielsen et al. [6], we choose $\mathbf{R} = \lambda \cdot \Sigma^{-1}$, where λ is a free parameter determining the regularization's weight.

Finally, note that by taking $\mathbf{S} = \mathbf{1}$ or $\mathbf{M} = \mathbf{1}$, the problem is reduced to sparse or multiplexed acquisition, respectively, with one exception: in our new formulation, it is not necessary to perform the demultiplexing step explicitly; instead, we directly obtain a denoised reconstruction.

5. Evaluation

We evaluate our approach on a material database of 12 pieces of leather and cloth, respectively. BTFs for the materials were captured by means of a semi-parallel setup consisting of 11 cameras, 198 LEDs, and a sample holder placed on a turntable which is rotated in increments of 30° during measurement.

For the purpose of numerical comparisons, we avoid inaccuracies due to registration and changes in scene geometry by obtaining both single-light and multiplexed images per turntable rotation. The resulting measured BTFs each consist of $198 \times 11 \times 12 = 26136$ HDR RGB images of size 128×128.

In a post-processing step, we use exposure bracking to obtain HDR images, apply de-mosaicing, rectify the images using 2D homographies, and finally apply color correction. For the ground truth data, this process takes a couple of hours. As it scales roughly linearly with the number of images (counting both exposure steps and number of pairs of incoming and outgoing light directions), the proposed method allows for significantly faster post-processing as well.

We randomly select one leather and two cloth BTFs as test materials. The remaining materials are used for training the models. All computations are performed using MATLAB 2015b on a 3 GHz dual-Xeon E5-2690 machine with 20 physical cores and 256 GB of RAM. We confirmed that with the given data, the amount of RAM required to run the algorithm was well below what was available and even suitable for less powerful desktop workstations.

As both our approach and the previous ones minimize the reconstruction error with respect to some metric, a direct purely numerical comparison between reconstruction results can hardly ever be fair, because one method would be compared against the other's metric, with respect to which it is supposed to be optimal. We thus compare the peak signal-to-noise ratio (PSNR) of renderings of the various reconstructions to renderings of the ground truth data, both under environmental lighting and directional lighting.

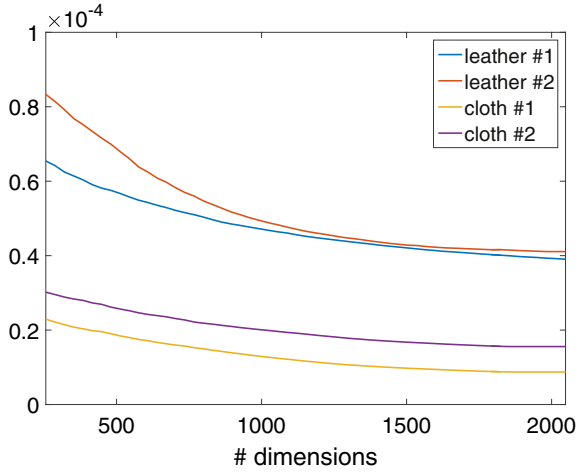
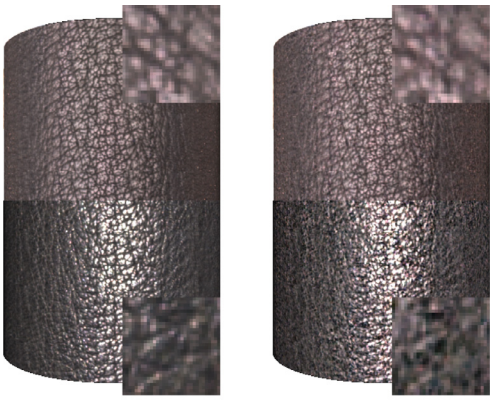
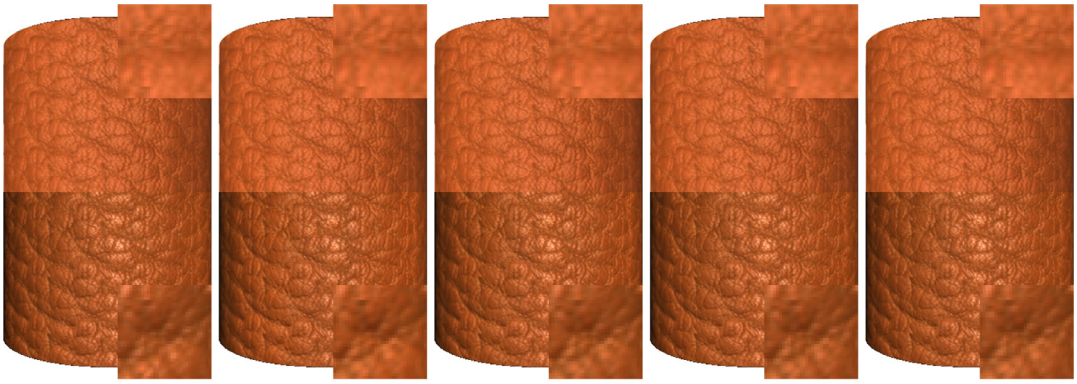


Fig. 5. Relative L^2 projection error, ours.



(a) Reference. (b) Projection, $d = 1024$.

Fig. 6. Evaluation of our basis for a challenging material exhibiting strong glossiness. The black artifacts are typical of insufficient range reduction and occur even when using $d = 1024$ basis vectors.



(a) Reference. (b) $d = 256$, PSNR 36.8 dB (c) $d = 512$, PSNR 39.9 dB (d) $d = 768$, PSNR 41.6 dB (e) $d = 1024$, PSNR 42.6 dB

Fig. 7. Basis projections for various numbers d of basis vectors, Nielsen et al. [6]. Top half: environmental lighting, bottom half: point light. Typically, $d = 512$ basis vectors suffice.

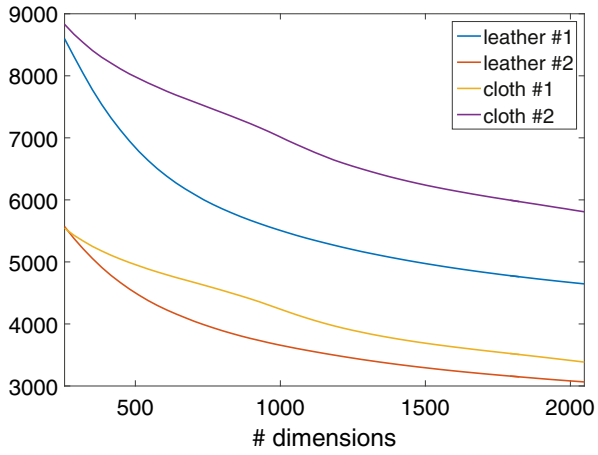


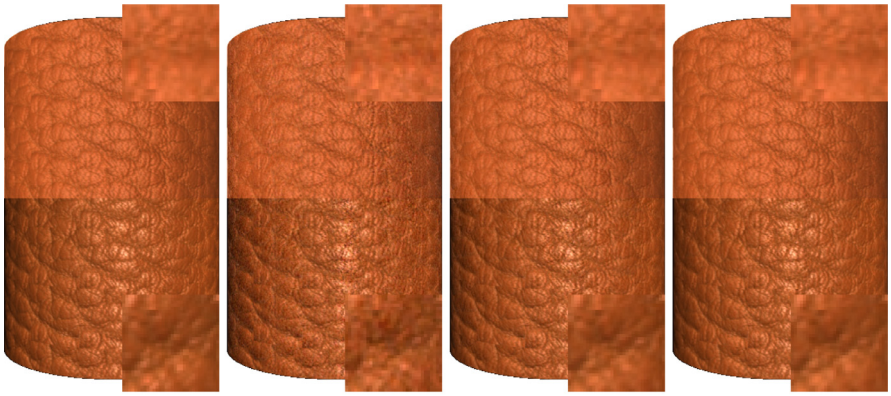
Fig. 8. Absolute L^2 projection error, Nielsen et al. [6].

5.1. Basis computation and performance

We first compute bases for each material separately as explained in Section 4, using a target dimensionality of $d_{\text{material}} = 256$, and the BTF matrix' element-wise inverse as weights. The algorithm converges quickly; we allowed for 15 iterations, reached after about two hours, which might have been excessive given that the relative changes in reconstruction errors we observed at that point were miniscule. The per-material bases are subsequently merged in negligible time using SVD as in Eq. (9). We obtain projections $\mathbf{U}^T\mathbf{B}_{\text{test}}$ for various numbers $d_{\text{class}} \in \{256, \ldots, 2048\}$ of columns of $\mathbf{U}$, which takes up to a couple of minutes, and evaluate the results numerically (cf. Fig. 5) and perceptually (cf. Fig. 4).

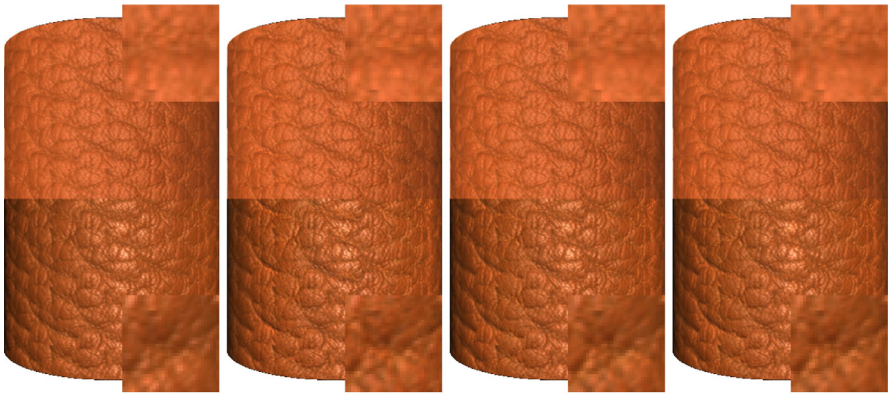
As can be seen in Fig. 5, contrary to what Matusik et al. [13] observed in the case of isotropic BRDFs, there is a clear plateau at very high dimensions at best; we thus choose $d_{\text{class}} = 1024$ by visual inspection. For comparison, we include reconstructions from projections to the basis proposed by Nielsen et al. [6], along with the absolute reconstruction error for projections to bases of several dimensions. To this end, we compute per-material bases following Nielsen et al. (cf. Section 3) and merge them afterwards to obtain a basis for the entire training set. Even more than with our bases, there is no plateau in the graph of reconstruction errors (cf. Fig. 8). Judging from visual inspection, we find that the resulting models are a bit more expressive than ours, in that even $d_{\text{class}} = 512$ basis vectors are typically sufficient (Fig. 7). In the following, we shall thus choose $d = 1024$ when evaluating our method, and $d = 512$ for the others.

While our models express most materials well, they appear to be limited to some extent by the presence of strong glossiness: as



(a) Reference. (b) $n_s = 100$, PSNR 30.0 dB (c) $n_s = 200$, PSNR 34.6 dB (d) $n_s = 400$, PSNR 37.7 db

Fig. 9. Reconstructions from sparse measurements for various numbers n_s of samples, ours. Top half: environmental lighting, bottom half: point light. $n_s >= 200$ samples are typically sufficient.



(a) Reference. (b) $n_s = 100$, PSNR 33.4 dB (c) $n_s = 200$, PSNR 35.8 dB (d) $n_s = 400$, PSNR 36.4 dB

Fig. 10. Reconstructions from sparse measurements for various numbers n_s of samples, Nielsen et al. [6]. Top half: environmental lighting, bottom half: point light. As with our method, $n_s >= 200$ samples are usually sufficient.

evident in Fig. 6, the reconstruction from the projection of such a material to our basis exhibits artifacts typical of insufficient reduction of dynamic range, demonstrated, e.g., by Menzel and Guthe [41] We believe this problem could be solved by appropriately adjusting the weights **W** in Eq. (5).

5.2. Reconstruction from sparse data

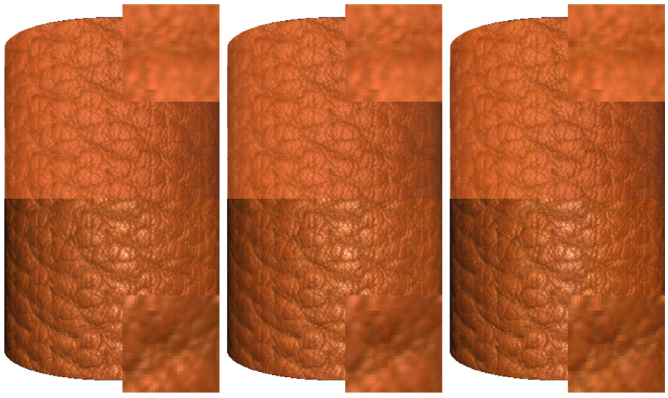
In our experiments, the sparse acquisition method proposed by Nielsen et al. [6] slightly but consistently outperformed its predecessor by den Brok et al. [5] on BTF data. We shall thus evaluate our method against Nielsen et al.'s only.

As explained in Section 4, we have to resort to Matusik et al.'s method [13] of determining the sparse samples to draw. Moreover, our acquisition setup is able to capture 11 images simultaneously. We thus further modified the sampling algorithm to only choose from pairs of light sources and turntable rotations and then always use all cameras available. That way, we obtain samplings for $n_s \in \{100, 200, 400\}$ pairs of light sources and turntable rotations, resulting in $\{1100, 2200, 4400\}$ samples or approximately $\{4.2, 8.4, 16.8\}$ per cent of 26,136 samples altogether. Due to our choice of sampling, acquisition time is reduced down to $\frac{n_s}{198 \cdot 12}$ times the reference data acquisition time.

We observe that our method is more sensitive to the choice of the regularization weight λ than that of Nielsen et al. We find that $\lambda = 0.01$ is a good choice for our approach, whereas we can follow Nielsen et al. and choose $\lambda = 40$ as they report.

With the given parameters, we produce reconstructions from sparse measurements as outlined in Sections 3 and 4. Overall, both methods appear to perform similarly well; in the case of leather #1, cf. Figs. 9 and 10, both methods produce visible artifacts for $n_s = 100$, with a slight numerical and perceptual advantage for Nielsen et al., whereas we slightly outperform Nielsen et al. in numerical terms for $n_s = 400$.

Note that these experiments depend on a number of tunable parameters (assumed dimensionality, regularization weight, color space transformation) that we might not have chosen optimally for either method; our main point here is that while our models appear to be a bit less expressive than the previous ones, they still can be used at least almost equally well when compared to the competing methods.



(a) Reference. (b) Ours, PSNR 37.6 dB (c) den Brok et al., PSNR 34.7 dB

Fig. 11. Reconstructions from illumination-multiplexed measurements. Top half: environmental lighting, bottom half: point light. Because our method does not require prior de-multiplexing, we do not have to deal with noise issues and therefore outperform the state-of-the-art on a number of materials, even though our model is slightly less expressive.

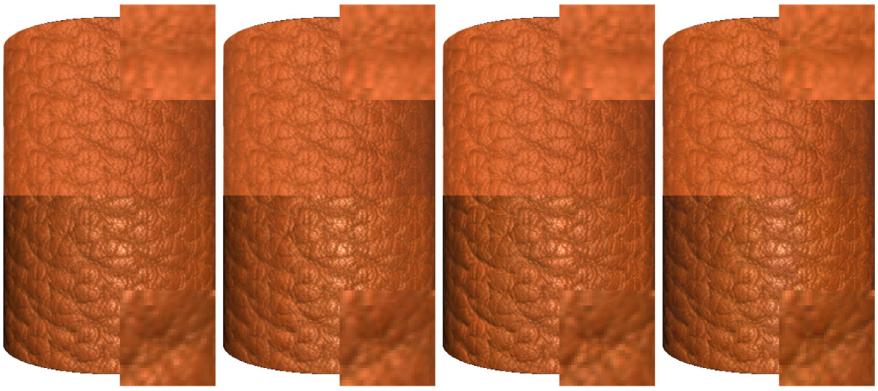
5.3. Demultiplexing

We choose n_p as small as possible according to den Brok et al. [7], i.e. greater than the number of light sources, such that a suitable set of Hadamard patterns exists, leading to $n_p = 199$.

We observe that a positive regularization weight $\lambda = 0.01$ improves slightly on reconstruction quality. As can be seen in Fig. 11, we are able to produce results of a quality similar to that of den Brok et al.'s method, which is the point we intend to make. Note that with our method, there is no need to de-multiplex the measured data first, which is how we are able to avoid the problems with noise caused by the de-multiplexing process.

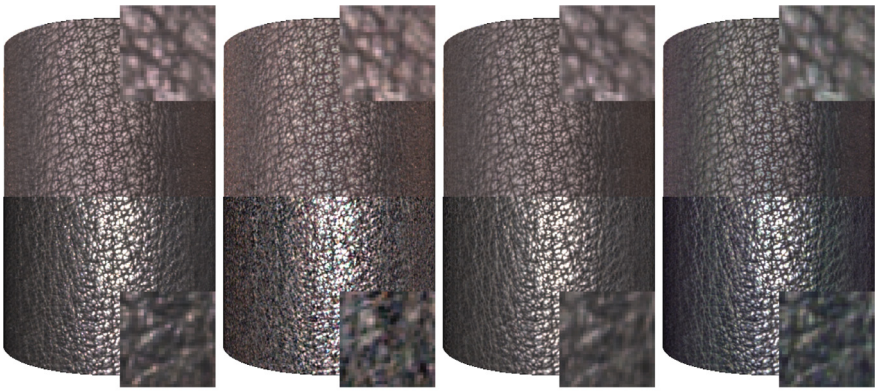
5.4. Sparse multiplexed acquisition

Finally, we present results for simultaneous sparse and multiplexed acquisition (cf. Figs. 12–15). For comparison, we include results for either method alone as proposed in previous work, along with reconstruction errors and approximate acquisition times. As sparsity, we choose $n_s = 400$, which provides a good compromise between acquisition speed-up and reconstruction quality. Over-



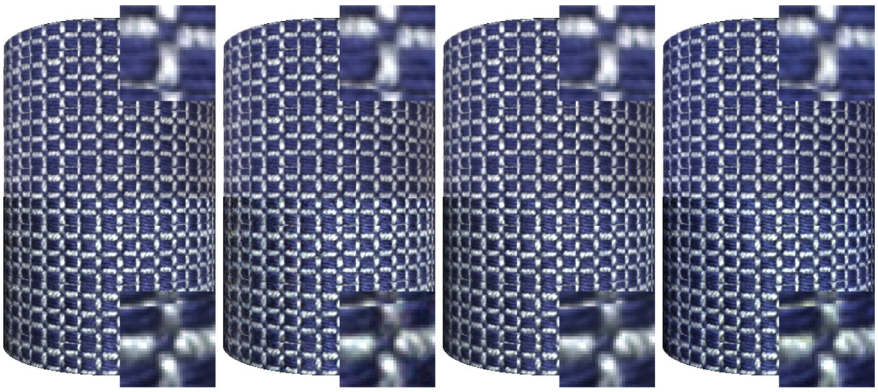
(a) Reference, $\Delta t \approx 22.8$ h | (b) Ours, PSNR 36.7 dB, $\Delta t \approx 12$ min | (c) Nielsen et al., PSNR 36.4 dB, $\Delta t \approx 4$ h | (d) den Brok et al., PSNR 34.8 dB, $\Delta t \approx 1.2$ h

Fig. 12. Acquisition times Δt and reconstructions from sparse, illumination-multiplexed measurements using $n_s = 400$ samples (ours), sparse, single-light measurements (Nielsen et al.), and full multiplexed measurements (den Brok et al.), leather #1. Top half: environmental lighting, bottom half: point light. Our method numerically outperforms both competing methods for this material, while allowing for greatly reduced acquisition times.



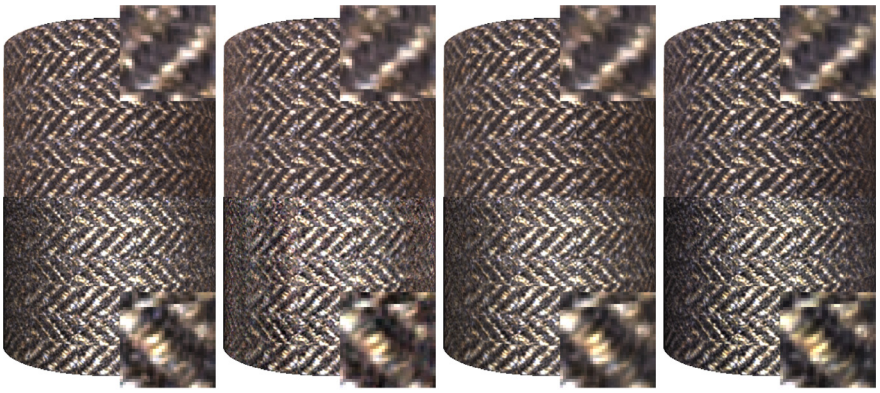
(a) Reference, $\Delta t \approx 20.6$ h | (b) Ours, PSNR 22.0 dB, $\Delta t \approx 16$ min | (c) Nielsen et al., PSNR 28.5 dB, $\Delta t \approx 3.4$ h | (d) den Brok et al., PSNR 25.7 dB, $\Delta t \approx 1.5$ h

Fig. 13. Acquisition times Δt and reconstructions from sparse, illumination-multiplexed measurements using $n_s = 400$ samples (ours), sparse, single-light measurements (Nielsen et al.), and full multiplexed measurements (den Brok et al.), leather #2. Top half: environmental lighting, bottom half: point light. We consider this a failure case of our method, likely due to our model's insufficiency to represent the chosen material well. However, the greatly reduced acquisition time might outweigh the reconstruction artifacts in some applications.



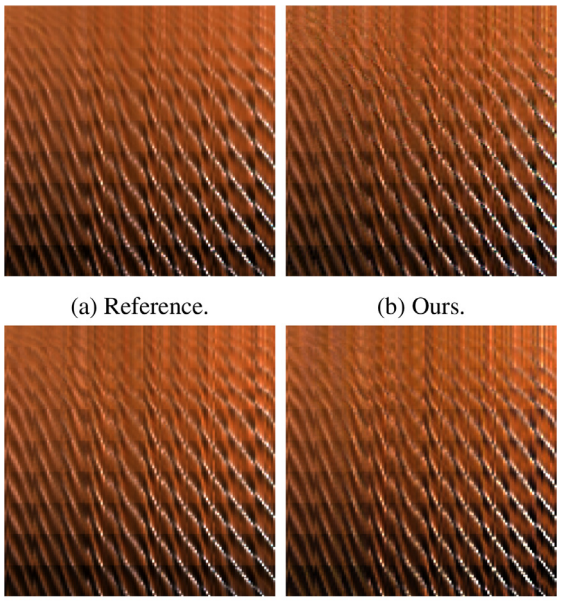
(a) Reference, $\Delta t \approx 10.8$ h | (b) Ours, PSNR 27.8 dB, $\Delta t \approx 16$ min | (c) Nielsen et al., PSNR 32.0 dB, $\Delta t \approx 1.8$ h | (d) den Brok et al., PSNR 28.1 dB, $\Delta t \approx 1.6$ h

Fig. 14. Acquisition times Δt and reconstructions from sparse, illumination-multiplexed measurements using $n_s = 400$ samples (ours), sparse, single-light measurements (Nielsen et al.), and full multiplexed measurements (den Brok et al.), cloth #1. Top half: environmental lighting, bottom half: point light. Reconstructions are very close visually, while Nielsen et al.'s method outperforms the others numerically. Again, acquisition times are greatly reduced with our method.



(a) Reference, $\Delta t \approx 11.6$ h | (b) Ours, PSNR 23.5 dB, $\Delta t \approx 22$ min | (c) Nielsen et al., PSNR 28.0 dB, $\Delta t \approx 2$ h | (d) den Brok et al., PSNR 25.7 dB, $\Delta t \approx 2.2$ h

Fig. 15. Acquisition times Δt and reconstructions from sparse, illumination-multiplexed measurements using $n_s = 400$ samples (ours), sparse, single-light measurements (Nielsen et al.), and full multiplexed measurements (den Brok et al.), cloth #2. Top half: environmental lighting, bottom half: point light. Our method is outperformed in numerical terms by the competing methods, but visually close and much faster to obtain.



(a) Reference. (b) Ours. (c) Nielsen et al. (d) den Brok et al.

Fig. 16. ABRDFs from various reconstructions of leather #1, for the same randomly selected texel. All methods exhibit minor artifacts in different areas, but overall, the reconstructions are close to the reference.

all, our method produces results comparable to those of previous methods, with the expected exception of leather #2, which our model so far does not represent well (cf. Fig. 13). This also holds on the level of ABRDFs, an example of which is shown in Fig. 16. However, in terms of acquisition times, our method significantly outperforms the state-of-the-art, which takes between 1.2 and 4 h, whereas we are able to produce reconstructions from sparse illumination-multiplexed measurements in 12–22 min.

6. Conclusion

To the best of our knowledge, we introduced the first technique for data-driven reflectance acquisition that combines both optimizing for the minimum set of view-light configurations and considering high dynamic range imaging in a single framework. For this purpose, we proposed a novel method for learning data-driven appearance models based on moving the dynamic range reduction from the data to the metric, where a novel basis for material BTFs is obtained using a metric computed via an efficient alternating-least squares algorithm and based on the relative L_2 error to reduce the impact of the reflectance data's dynamic range. We demonstrated that like previously studied bases, our basis can be used for the purpose of sparse acquisition and de-noising of data obtained via multiplexed illumination, but additionally allows for these approaches to be combined to the effect of greatly reduced acquisition times in the order of minutes, which is orders of magnitude faster than traditional image-by-image measurements, and several to many times faster than previous approaches, while at the same time maintaining reconstruction results of comparable quality.

As of now, our basis is slightly less expressive than bases obtained using a logarithmic scaling of the data, in that the dimension of its span is larger. We believe that future work might further reduce dimensionality by choosing appropriate admissible transformations and weights such as the ones proposed by Bagher et al. [37] in the context of compressing isotropic BRDFs using tensor factorization. This might also solve or mitigate the problem we observed with the rather challenging leather #2 (cf. Fig. 13), which we believe is due to its rather high degree of glossiness. Moreover, the range of materials we evaluated our method on is rather limited; it would be interesting to investigate its performance on other categories of materials, and whether there is a model that works for several (or all) classes.

References

[1] Nicodemus FE, Richmond JC, Hsia JJ, Ginsberg IW, Limperis T. Geometrical considerations and nomenclature for reflectance. National Bureau of Standards Monograph #160; U.S. Department of Commerce; 1977.
[2] Aittala M, Weyrich T, Lehtinen J. Two-shot SVBRDF capture for stationary materials. ACM Trans Graph 2015;34(4):110:1–110:13. doi:10.1145/2766967.
[3] Dana KJ, van Ginneken B, Nayar SK, Koenderink JJ. Reflectance and texture of real-world surfaces. ACM Trans Graph 1999;18(1):1–34. doi:10.1145/300776.300778.
[4] Schwartz C, Sarlette R, Weinmann M, Rump M, Klein R. Design and implementation of practical bidirectional texture function measurement devices focusing on the developments at the university of bonn. Sensors 2014;14(5). doi:10.3390/s140507753.
[5] den Brok D, Steinhausen HC, Hullin MB, Klein R. Patch-based sparse reconstruction of material BTFs. J WSCG 2014;22(2):83–90.
[6] Nielsen JB, Jensen HW, Ramamoorthi R. On optimal, minimal BRDF sampling for reflectance acquisition. ACM Trans Graph 2015;34(6):186:1–186:11. doi:10.1145/2816795.2818085.

[7] den Brok D, Steinhausen HC, Klein R. Fast multiplexed acquisition of high-dynamic-range material appearance. In: Bommes D, Ritschel T, Schultz T, editors. Vision, modeling & visualization. Aachen, Germany: Eurographics Association; 2015. p. 151–8. doi:10.2312/vmv.20151270.
[8] Müller G, Meseth J, Sattler M, Sarlette R, Klein R. Acquisition, synthesis and rendering of bidirectional texture functions. In: Eurographics 2004, state of the art reports; 2004. p. 69–94.
[9] Filip J, Haindl M. Bidirectional texture function modeling: a state of the art survey. IEEE Trans Pattern Anal Mach Intell 2009;31(11):1921–40.
[10] Haindl M, Filip J. Visual texture: accurate material appearance measurement, representation and modeling. Advances in computer vision and pattern recognition. London: Springer-Verlag; 2013.
[11] Weinmann M, Klein R. Advances in geometry and reflectance acquisition (course notes). In: Proceedings of the SIGGRAPH Asia 2015 courses. New York, NY, USA: ACM; 2015. p. 1:1–1:71. doi:10.1145/2818143.2818165.
[12] Wong T-T, Heng P-A, Or S-H, Ng W-Y. Image-based rendering with controllable illumination. In: Proceedings of the eurographics workshop on rendering techniques '97. London, UK, UK: Springer-Verlag; 1997. p. 13–22. ISBN 3-211-83001-4, http://dl.acm.org/citation.cfm?id=647651.731971
[13] Matusik W, Pfister H, Brand M, McMillan L. A data-driven reflectance model. ACM Trans Graph 2003a;22(3):759–69. doi:10.1145/882262.882343.
[14] Matusik W, Pfister H, Brand M, McMillan L. Efficient isotropic BRDF measurement. In: Proceedings of the fourteenth eurographics workshop on rendering. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association; 2003b. p. 241–7. EGRW '03
[15] Koudelka ML, Magda S, Belhumeur PN, Kriegman DJ. Acquisition, compression, and synthesis of bidirectional texture functions. In: Proceedings of the ICCV workshop on texture analysis and synthesis; 2003.
[16] Dong Y, Wang J, Tong X, Snyder J, Lan Y, Ben-Ezra M, et al. Manifold bootstrapping for SVBRDF capture. ACM Trans Graph 2010;29(4):98:1–98:10. doi:10.1145/1778765.1778835.
[17] Xu Z, Nielsen J, Yu J, Jensen H, Ramamoorthi R. Minimal BRDF sampling for two-shot near-field reflectance acquisition. ACM Trans Graph 2016;35(6). doi:10.1145/2980179.2982396.
[18] Yu J, Xu Z, Mannino M, Jensen HW, Ramamoorthi R. Sparse Sampling for Image-Based SVBRDF Acquisition. In: Klein R, Rushmeier H, editors. Workshop on material appearance modeling. The Eurographics Association; 2016.
[19] Zhou Z, Chen G, Dong Y, Wipf D, Yu Y, Snyder J, et al. Sparse-as-possible SVBRDF acquisition. ACM Trans Graph 2016;35(6):189:1–189:12. doi:10.1145/2980179.2980247.
[20] Peers P, Mahajan DK, Lamond B, Ghosh A, Matusik W, Ramamoorthi R, et al. Compressive light transport sensing. ACM Trans Graph 2009;28(1):3:1–3:18. doi:10.1145/1477926.1477929.
[21] Marwah K, Wetzstein G, Bando Y, Raskar R. Compressive light field photography using overcomplete dictionaries and optimized projections. ACM Trans Graph 2013;32(4):46:1–46:12. doi:10.1145/2461912.2461914.
[22] Miandji E, Kronander J, Unger J. Compressive image reconstruction in reduced union of subspaces. In: Proceedings of the eurographics 2015; 2015.
[23] Davis A, Levoy M, Durand F. Unstructured light fields. Comput Graph Forum 2012;31(2pt1):305–14. doi:10.1111/j.1467-8659.2012.03009.x.
[24] Thanikachalam N, Baboulaz L, Firmenich D, Süsstrunk S, Vetterli M. Scene relighting with smartphones. IEEE Trans Comput Imaging 2016.
[25] Aittala M, Weyrich T, Lehtinen J. Practical SVBRDF capture in the frequency domain. ACM Trans Graph 2013;32(4):110:1–110:12. doi:10.1145/2461912.2461978.
[26] Aittala M, Aila T, Lehtinen J. Reflectance modeling by neural texture synthesis. ACM Trans Graph 2016;35(4):65:1–65:13. doi:10.1145/2897824.2925917.
[27] Harwit M, Sloane N. Hadamard transform optics. Academic Press; 1979.
[28] Wenger A, Gardner A, Tchou C, Unger J, Hawkins T, Debevec P. Performance relighting and reflectance transformation with time-multiplexed illumination. ACM Trans Graph 2005;24(3):756–64. doi:10.1145/1073204.1073258.
[29] Schechner YY, Nayar SK, Belhumeur PN. Multiplexing for optimal lighting. IEEE Trans Pattern Anal Mach Intell 2007;29(8):1339–54. doi:10.1109/TPAMI.2007.1151.
[30] Ratner N, Schechner Y. Illumination multiplexing within fundamental limits. In: Proceedings of the computer vision and pattern recognition, 2007. CVPR '07. IEEE conference on; 2007. p. 1–8. doi:10.1109/CVPR.2007.383162.
[31] Park JI, Lee MH, Grossberg MD, Nayar SK. Multispectral imaging using multiplexed illumination. In: Proceedings of the international conference on computer vision; 2007. p. 1–8. doi:10.1109/ICCV.2007.4409090.
[32] Liang C-K, Lin T-H, Wong B-Y, Liu C, Chen HH. Programmable aperture photography: multiplexed light field acquisition. ACM Trans Graph 2008;27(3):55:1–55:10. doi:10.1145/1360612.1360654.
[33] Mitra K, Cossairt O, Veeraraghavan A. Can we beat Hadamard multiplexing? data driven design and analysis for computational imaging systems. In: Proceedings of the 2014 IEEE international conference on computational photography (ICCP),; 2014. p. 1–9. doi:10.1109/ICCPHOT.2014.6831800.
[34] Miyagawa I, Taniguchi Y. Dense light transport for relighting computation using orthogonal illumination based on Walsh–Hadamard matrix. IEICE Trans Inf Syst 2016;E99-D(4):1038–51.
[35] Vavra R, Filip J. Minimal sampling for effective acquisition of anisotropic BRDFs. Comput Graph Forum 2016(7):299–309. (PACIFIC GRAPHICS 2016) doi:10.1111/cgf.13027.
[36] Ruiters R, Schwartz C, Klein R. Data driven surface reflectance from sparse and irregular samples. Comput Graph Forum 2012;31(2):315–24. (Proc of Eurographics).
[37] Bagher MM, Snyder J, Nowrouzezahrai D. A non-parametric factor microfacet model for isotropic BRDFs. ACM Trans Graph 2016;35(5):159:1–159:16. doi:10.1145/2907941.
[38] Hitchcock FL. The expression of a tensor or a polyadic as a sum of products. J Math Phys 1927;6(1–4):164–89. doi:10.1002/sapm192761164.
[39] Hall P, Marshall D, Martin R. Merging and splitting eigenspace models. IEEE Trans Pattern Anal Mach Intell 2000;22(9):1042–9. doi:10.1109/34.877525.
[40] Rusinkiewicz SM. Vienna: Springer Vienna; 1998. p. 11–22.
[41] Menzel N, Guthe M. High dynamic range preserving compression of light fields and reflectance fields. In: Proceedings of the 5th international conference on computer graphics, virtual reality, visualisation and interaction in Africa. AFRIGRAPH '07. New York, NY, USA: ACM; 2007. p. 71–6. doi:10.1145/1294685.1294697.

APPENDIX E

Per-image super-resolution for material BTFs

©IEEE 2020. Reprinted, with permission, from

D. den Brok, S. Merzbach, M. Weinmann and R. Klein, "Per-Image Super-Resolution for Material BTFs," 2020 IEEE International Conference on Computational Photography (ICCP), 2020, pp. 1-10, doi: 10.1109/ICCP48838.2020.9105256.

Per-Image Super-Resolution for Material BTFs

Dennis den Brok, Sebastian Merzbach, Michael Weinmann, and Reinhard Klein

Abstract—Image-based appearance measurements are fundamentally limited in spatial resolution by the acquisition hardware. Due to the ever-increasing resolution of displaying hardware, high-resolution representations of digital material appearance are desireable for authentic renderings. In the present paper, we demonstrate that high-resolution bidirectional texture functions (BTFs) for materials can be obtained from low-resolution measurements using single-image convolutional neural network (CNN) architectures for image super-resolution. In particular, we show that this approach works for high-dynamic-range data and produces consistent BTFs, even though it operates on an image-by-image basis. Moreover, the CNN can be trained on down-sampled measured data, therefore no high-resolution ground-truth data, which would be difficult to obtain, is necessary. We train and test our method's performance on a large-scale BTF database and evaluate against the current state-of-the-art in BTF super-resolution, finding superior performance.

Index Terms—Super-Resolution, Appearance Capture, Deep Learning

1 Introduction

Super-resolution, i.e. artificially increasing the sampling rate of a given measuring system, is a problem faced in disciplines as diverse as astronomy and biology. In computer vision and graphics, it typically deals with increasing the image resolution of imaging systems of all kinds, from smartphone cameras to light-field cameras, for static data as well as dynamic. There is an abundance of practical solutions for many of these applications. In contrast, imaging systems of importance to rendering applications, such as acquisition devices for material appearance, have seen a lot less attention. But demand has increased driven by rapid developments in display technology, and while image-space super-resolution for Lambertian reflectance has been studied and understood for a long time, more comprehensive representations like spatially-varying bidirectional reflectance distribution functions (SVBRDFs) and bidirectional texture functions (BTFs) have started to draw interest only recently. This case is a lot more difficult to deal with, as even though typical measurements consist of hundreds or thousands of images with what seems like exploitable per-texel redundancies, the fact that the material's reflection at each texel can often only be described by complex BRDFs renders traditional multi-view super-resolution approaches inapplicable. Consequently, the current state-of-the-art method for BTF super-resolution by Dong et al. [1] is based on classical single-image super-resolution, albeit applied to eigentextures for a significant performance increase.

In the present paper, we take things a step further and introduce what we believe to be the first super-resolution algorithm for non-Lambertian spatially-varying material appearance based on modern deep learning techniques for single-image super-resolution. Our goal in this work has not been to find the existing method to provide the best possible results for our use cases, which would have involved training and comparing an abundance of network architectures. Instead, we try to answer a number of questions related to the problem at hand:

- Are typical single-image super-resolution networks suitable for the high dynamic range (HDR) radiance maps our setup produces?
- Do we need ground-truth data in the target resolution, which may be difficult to obtain, or is the problem scale-invariant to some extent?
- Does the image-by-image approach lead to visible artefacts in the resulting material representations?

To do so, we construct a simple convolutional neural network (CNN) by extending one of the pioneering works with current best pactices. We train and test our network on a database of down-sampled HDR radiance maps from our material measurements. From the resulting high-resolution test images, we produce high-resolution BTFs for which we provide both a ground-truth comparison and an evaluation against the state-of-the-art. In the latter case, we are able to demonstrate significantly improved reconstruction quality.

- *All authors are with the Institute of Computer Science II, University of Bonn, Germany.*
  E-mail: denbrok@cs.uni-bonn.de

2 Related Work

There is a wealth of literature on the general subject of super-resolution. We briefly review some of the publications more relevant to our work and refer to surveys for further reference.

Single-Image Super-Resolution

Approaches for restoring high-resolution representations from a single low-resolution images can be categorized into interpolation-based, reconstruction-based and learning-based methods. Interpolation-based techniques such as linear interpolation, bicubic interpolation or Lanczos resampling [2] offer the possibility of a fast and simple upsampling at the cost of limited accuracy due to remaining blurring artifacts. In contrast, the use of additional prior knowledge allows reconstruction-based approaches (e.g. [3],

©2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes,creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

[4], [5], [6]) to be more flexible and even reproduce high-frequency structures. However, these techniques are time-consuming and suffer from a decreasing performance for increasing scales. Learning-based approaches exploit the statistical relationships between low-resolution images and their corresponding high-resolution exemplars and have been proven to allow results of high fidelity at fast computation. Classical learning-based approaches comprise the use of Markov random fields [7], [8] and sparse coding [9], [10], [11], [12], [13], [14], [15]. In recent years, these techniques have been outperformed by deep learning approaches [16], [17], [18], [19]. The Super-Resolution Convolutional Neural Network (SRCNN) is based on an initial traditional bicubic upsampling of the image resolution which is followed by a CNN-based refinement. Only forcing the network to perform the refinement without the upsampling facilitates the task and the restriction to convolutional layers allows to handle input images of arbitrary size. The involved CNN consists of three convolutional layers that, except the last convolutional layer, are followed by ReLU layers and take the role of patch/feature extraction, nonlinear mapping of the feature maps to a high-dimensional vector and reconstruction by aggregating the feature maps. Other approaches extend this idea by using different architectures and learning strategies. However, performing most of the operations in the high-resolution representation comes at the cost of a high computational burden. To reduce the latter and allow for a faster training and inference, several techniques perform feature extraction is performed in the low-resolution image directly and afterwards involve end-to-end learnable upsampling layers [17], [20]. Respective variants of these techniques involve different learnable upsampling layers, structures and learning strategies. As performing the upsampling in a single step is difficult for large scaling factors, further developments aim at predicting the output in multiple steps and, hence, in a progressive manner [21], [22], [23], [24]. Further frameworks are based on generative adversarial networks (GANs) [18], [19], [25], [26], [27], [28] where a generator is trained to created super-resolution images that cannot be distinguished from real high-resolution images by a discriminator. Wu et al. [27] additionally used a perceptual loss. For more detailed discussions, we refer to recent surveys [29], [30], [31].

BTF Super-Resolution

Dong et al. [1] address BTF image super-resolution (SR) based on factorizing the captured low-resolution BTF data into eigen-textures and eigen-apparent bidirectional reflectance distribution functions (eigen-ABRDFs) via singular value decomposition (SVD). This allows improving the resolution of the eigen-textures using classical image super-resolution techniques and the final high-resolution BTF results from the combination of high-resolution intrinsic textures and low-resolution eigen-ABRDFs. Zhang et al. [32] exploit the combination of a GAN [33] and the perception-driven texture generation model (PDTG) [34] to learn material appearance under different illumination and viewing directions. Furthermore, Rainer et al. [35] presented an autoencoder-based approach for BTF representation, where BTF texels are compressed to latent codes by the encoder and the decoder is used to reproduce material appearance under guidance by specified light and view direction.

3 Background

Our goal is an algorithm capable of producing BTFs with high texture resolution from such with low texture resolution. There are a number of natural choices for what the algorithm's input should be, precisely. In order to be able to justify our particular choice, we first briefly introduce the notion of BTFs and explain how they can be represented digitally. Subsequently, we give an overview of the process from physical material sample to digitally represented BTF, and finally, as their method builds upon some of the techniques explained in this section, we briefly cover the prior art by Dong et al. [1]

3.1 Bidirectional Texture Functions

Consider the rendering equation

$$L_o(\mathbf{x}, \omega_i, \omega_o) = \int_{\Omega_i} f_r(\mathbf{x}, \omega_i, \omega_o) L_i(\mathbf{x}, \omega_i, \omega_o) d\omega_i \quad (1)$$

Most commonly, f_r is assumed to be a spatially-varying BRDF, i.e. it adheres to conservation of energy and Helmholtz reciprocity, and $\mathbf{x}$ is defined with respect to the given surface geometry. For BTFs, none of these assumptions are required to hold, which allows them to account for a number of effects not local to a texel, such as parallax and meso-scale inter-reflections. The downside is that they are notoriously difficult to model in a meaningful way, which is why most BTF representations are discrete and image-based. One of the main applications is arguably rendering, particularly in real-time. Due to the way BTFs are typically evaluated at given coordinates $(\mathbf{x}, \omega_i, \omega_o)$, it is convenient to assume the set of sampled pairs (ω_i, ω_o) to be a Cartesian product $\mathcal{L} \times \mathcal{V}$ of sets of light and view directions, respectively. In that case, a BTF can be represented as a tensor $\mathbf{B} \in \mathbf{R}^{c \times n_l \times n_v \times w \times h}$, where c is the number of color channels, n_l, n_v the number of sampled light and view directions, and w, h the size of the bidirectional textures, or more commonly as a matrix $\mathbf{B} \in \mathbf{R}^{n_l \cdot n_v \times c \cdot w \cdot h}$. This is the desired final form for our algorithm's output.

3.2 BTF Compression

In most practical scenarios, $\mathcal{L}$ and $\mathcal{V}$ need to be sufficiently dense as to avoid interpolation artifacts when the BTF is used in rendering. A typical order of magnitude is $|\mathcal{L} \times \mathcal{V}| \propto 10^5$, which leads to impractical amounts of data. It is known, however, that BTFs in the matrix representation outlined in Section 3.1 admit a low-rank factorization

$$f(\mathbf{B}) \approx \mathbf{U} \cdot \mathbf{V}, \quad (2)$$

where $\mathbf{U} \in \mathbf{R}^{n_l \cdot n_v \times k}$, $\mathbf{V} \in \mathbf{R}^{k \times c \cdot w \cdot h}$ are matrices of rank $k \ll n_l \cdot n_v$ [36], which are usually obtained by singular value decomposition (SVD), and f an optional function modifying the effective metric used when computing the factorization, commonly used to deal with the wide dynamic range of BTF data. Common values for k are in the range 100–500.

Dong et al. [1] exploit this factorization to achieve fast and noise-resistant BTF super-resolution by operating per-eigentexture, i.e. per row of the matrix $\mathbf{V}$, only. One caveat is that the algorithm relies on the eigentextures to be image-like, i.e. $f = \mathrm{id}$, and therefore likely works best for materials with low-dynamic-range BTF.

3.3 Measuring BTFs

Clearly, it is impractial to sample a material's BTF at precisely the desired coordinates, as the directions of incoming and outgoing light vary across the material surface. Virtually every available acquisition device therefore produces a set of images of the material as seen and lit from different angles as an intermediate step. In principle, the data acquired this way already constitutes a sampling of the material's BTF, but for the purpose of storage and rendering efficiency, it is convenient to represent the data as a matrix of the shape outlined in Section 3.1. This involves a number of transformations which alter the image data significantly:

First, in order to obtain proper textures, a homographical transformation is applied image-by-image which warps the region of interest on the material surface from a polygonal shape in image space to a rectangular shape of size $w \times h$. As a result, textures obtained from images taken by cameras seeing the material surface from a grazing angle contain information with much lower frequencies than those from images as seen from a frontal parallel view.

Subsequently, the textures can be arranged in a matrix representation of the described shape, but the columns, which correspond to per-texel BRDFs, do not yet correspond to the desired sampling $\mathcal{L} \times \mathcal{V}$. The BRDFs are thus re-sampled accordingly, typically by means of multilinear interpolation, which further smoothes the textures.

3.4 Prior Art

The only algorithm for BTF super-resolution to date has been proposed by Dong et al. [1] Similarly to ours it exploits single-image super-resolution, but applied to eigentextures instead of actual images, which provides both a significant speed-up and improved robustness against noise in the data. Contrary to the present paper, they chose a classical approach based on deconvolution with a smoothness prior:

Starting at the image formation model

$$\mathbf{y} = \mathbf{DPx} + \mathbf{n}, \tag{3}$$

where $\mathbf{D} \in \mathbf{R}^{w_{\mathrm{LR}} \cdot h_{\mathrm{LR}} \times w_{\mathrm{HR}} \cdot h_{\mathrm{HR}}}$ is a downsampling operator, $\mathbf{P} \in \mathbf{R}^{w_{\mathrm{HR}} \cdot h_{\mathrm{HR}} \times w_{\mathrm{HR}} \cdot h_{\mathrm{HR}}}$ a blur matrix accounting for the camera's point spread function (PSF), $\mathbf{n} \in \mathbf{R}^{w_{\mathrm{LR}} \cdot h_{\mathrm{LR}}}$ a noise vector, and $\mathbf{x} \in \mathbf{R}^{w_{\mathrm{HR}} \cdot h_{\mathrm{HR}}}$ and $\mathbf{y} \in \mathbf{R}^{w_{\mathrm{LR}} \cdot h_{\mathrm{LR}}}$ high- and low-resolution images, respectively, with $w_{\mathrm{HR}} > w_{\mathrm{LR}}$ and $h_{\mathrm{HR}} > h_{\mathrm{LR}}$, one arrives at the single-image super-resolution optimization problem

$$\hat{\mathbf{x}} = \mathrm{argmin}_{\mathbf{x}} \|\mathbf{DPx} - y\|_2^2 + \alpha\phi(\mathbf{x}), \tag{4}$$

where $\phi\colon \mathbf{R}^{w_{\mathrm{HR}} \cdot h_{\mathrm{HR}}} \longrightarrow \mathbf{R}$ is some regularizing function with weight α. Dong et al. [1] chose $\phi(\mathbf{x}) = \|\mathbf{Lx}\|_2^2$, where $\mathbf{L}$ is the 2D Laplacian operator. The resulting optimization problem can be solved very efficiently using the regularized normal equations. However, it is known to be sensitive to the presence of noise. Moreover, it has to be applied to each of the BTF's bidirectional textures individually. Dong et al. [1] thus enhance both efficiency and robustness of their optimization problem by operating on the individual eigentextures $\mathbf{V}$ of the factorized low-resolution BTF $\mathbf{B} \approx \mathbf{UV}$.

Note that $\mathbf{U}$ is computed on the low-resolution measured data, which may limit the achievable amount of detail to some extent. Note also that the measured PSF does not directly apply to the bidirectional textures and, by extension, to the eigentextures of the BTF, because they have undergone warping to account for perspective transformation; see Section 3.3. We are not aware of whether Dong et al. took the latter into consideration, but it did not seem to noticeably influence the quality of the results in our experiments.

Arguably, the speed improvement, though significant, is not the algorithm's primary feature, as the deconvolution is very fast even on an full BTF measurement; nevertheless we shall provide a suggestion how to reach comparable performance with our image-by-image approach.

4 Proposed Method

It should be clear from Section 3 that there are various natural ways to attack the problem of image-by-image super-resolution for BTFs. In the following we shall give a rationale for our particular design choices and a detailed description of the chosen network architecture.

4.1 Design Choices

First of all – why did we choose an image-by-image approach?

Recall that measured material BTFs are of the shape $\mathbf{B} \in \mathbf{R}^{n_l \cdot n_v \times c \cdot w \cdot h}$ or similar, with $n_l \cdot n_v \propto 10^5$ (cf. Section 3.1). This brings with it a number of problems. No super-resolution algorithm we are aware of is capable of dealing with anywhere near the necessary amount of $n_l \cdot n_v$ channels. Moreover, the BTF database at our disposal consists of only 24 BTFs with small $w \times h = 128 \times 128$, which is likely insufficient for learning proper parameters even when dividing the available BTFs up into patches. The number of channels can be decreased significantly by compressing the BTFs and operating on the eigentextures $\mathbf{V}$, similar to Dong et al. However, the high-resolution eigentexures thus obtained are parameters for a model $\mathbf{U}$ of eigen-BRDFs computed on the low-resolution BTF (as the high-resolution BTF is not available), which may limit the effectiveness of the up-sampling network. Nevertheless, we conducted experiments with this approach and indeed found it to perform worse than the state-of-the-art, likely due to the still high number of channels (typically 128–256) or the low number of exemplars to learn from.

Secondly, now that we motivated an image-by-image approach, the choice remains between the types of image data that occur in the post-processing process outlined in Section 3.3: eigentextures, bidirectional textures from the re-sampled BTF, warped regions of interest, or simply the HDR radiance maps before any further post-processing.

We see no obvious theoretical *a priori* reason why any should perform significantly better than the others. By experimentation we found that using HDR radiance maps in

fact works best by quite a margin. We assume that the filters learned from this kind of images are the ones enabling the highest-frequency reconstructions while not being too general: for images corresponding to warped regions of interest, the majority of the images are low-frequency because of the large amount of warping for low-angle views, and bidirectional textures have undergone even further interpolations. Conversely, the range of frequencies is extremely wide in the case of eigentextures, with the eigentexture corresponding to the largest eigenvalue being similar to a mean texture and the ones corresponding to very small eigenvalues being almost indistinguishable from noise. Please note, however, that this is only a hypothesis which we did not verify experimentally.

4.2 Network Architecture

The proposed network is a convolutional neural network (CNN) with skip connections, inspired by the work by Shi et al. [20] We follow their suggestions closely, but modify them in a few places according to current good practices. Their network consists of a L layers, the first $L-1$ of which are consecutive 2D convolutional layers, followed by a single sub-pixel convolution layer:

$$\begin{aligned} f^{(1)}(\mathbf{I}_{\mathrm{LR}}) &= \phi(W_1 * \mathbf{I}_{\mathrm{LR}} + b_1), \\ f^{(l)}(\mathbf{I}_{\mathrm{LR}}) &= \phi(W_l * f^{(l-1)}(\mathbf{I}_{\mathrm{LR}}) + b_l), \\ f^{(L)}(\mathbf{I}_{\mathrm{LR}}) &= \mathcal{PS}(W_L * f^{(L-1)}(\mathbf{I}_{\mathrm{LR}}) + b_L), \end{aligned} \tag{5}$$

where $\mathbf{I}_{\mathrm{LR}}$ is the low-resolution input image, W_l and b_l, $1 \leq l \leq L$, are the weights and biases of the corresponding layers, respectively, ϕ is the activation function (the same for all except the last layer), and $\mathcal{PS}$ is a periodic shuffling operator. In this scenario, $b_l \in \mathbf{R}^{n_l}$ is a vector of length n_l, and $W_l \in \mathbf{R}^a$ an 2D convolution tensor of size $n_{l-1} \times n_l \times k_l \times k_l$, where n_l is the number of features at layer l, $n_0 = 1$, and k_l the corresponding kernel size.

The original network did not perform very well in our scenario in terms of upsampling quality, which is why we introduced some modifications: Following Tong et al. [37] we added a *skip connection*:

$$f^{(L-1)}(\mathbf{I}_{\mathrm{LR}}) = \phi(W_{L-1} * \begin{pmatrix} f^{(1)}(\mathbf{I}_{\mathrm{LR}}) \\ f^{(L-2)}(\mathbf{I}_{\mathrm{LR}}) \end{pmatrix} + b_{L-1}), \tag{6}$$

which allows for combination of low- and high-level features, resulting in enhanced sharpness. See Figure 1 for a schematic of the resulting architecture. Moreover, we added *batch normalization* to deal with internal covariate shift, allowing for higher learning rates. [38]

We chose this particular architecture for several reasons: it is easy to understand, fast and efficient, in particular due to fact that the upsampling happens at the last layer only, and it does not make any assumptions about the desired output. The latter may seem like a disadvantage; however, in our material BTF scenario, we are interested more in faithful reconstruction in the sense of visual similarity rather than, e.g., an extremely sharp, plausible result which only remotely resembles the material sample. See Figure 4.2 for a reconstruction obtained using a recent, pre-trained state-of-the-art network [39] which may be sharper than what our network produces, but does not look like the original material sample. In particular, if this algorithm were applied on a per-image basis to an entire BTF, it would probably lead to visible artefacts.

4.3 Performance Considerations

Contrary to the work by Dong et al., our network does not seem to work well with eigentextures as input images, as hinted at in Section 4.1. However, once the network is trained, it is still not necessary to upsample *all* measured images. According to den Brok et al. [40], BRDF reconstruction from sparse measurements as proposed by Nielsen et al. [41] also applies to spatially-varying material appearance via

$$\mathbf{V} = (\mathbf{U}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\mathbf{S}\mathbf{U} + \mathbf{\Gamma}^{\mathrm{T}}\mathbf{\Gamma})^{-1}\mathbf{U}^{\mathrm{T}}(\mathbf{S}\mathbf{B}), \tag{7}$$

where $\mathbf{S}$ is a subsampling matrix which essentially selects rows, i.e. bidirectional textures, from the full BTF $\mathbf{B}$, and $\mathbf{\Gamma}$ is a diagonal matrix of weights giving priority to the eigen-BRDFs corresponding to the largest singular values. By leveraging this method we can therefore reduce the amount of images that need to be upsampled significantly, if desired.

Note there are two natural choices for $\mathbf{U}$ here, which represent a trade-off between speed and reconstruction quality: we may compute a basis $\mathbf{U}_{\mathrm{LR}}$ on the low-resolution input BTF, where, contrary to Dong et al., we have the advantage that we can apply non-linear transformations like the logarithm for dynamic-range reduction and thereby obtain more adequate and concise bases (cf. [42]). That way, we only have to upsample approximately the same amount of images as Dong et al., albeit putting an upper limit on the sharpness of the reconstruction because $\mathbf{U}_{\mathrm{LR}}$ was computed on low-resolution data. The alternative is to first produce a database of high-resolution BTFs using our network, and to use that database to obtain a basis $\mathbf{U}_{\mathrm{HR}}$ potentially allowing for more high-frequency reconstructions. As $\mathrm{rank}(\mathbf{U}_{\mathrm{HR}}) > \mathrm{rank}(\mathbf{U}_{\mathrm{LR}})$, more sample images need to be upsampled prior to sparse reconstruction, which reduces reconstruction speed.

Either way, this approach comes with a penalty on reconstruction quality, which has been studied before. [43] In our experiments, we therefore do *not* use this performance enhancement and reconstruct full BTFs instead.

5 Experimental Results

5.1 Experimental Setup

We evaluate the proposed algorithm on a database of 24 material samples, 12 of which are leathers and fabrics, respectively (cf. Table 1).

5.1.1 Acquisition Setup

Our acquisition device consists of 11 industrial-grade cameras and 198 LED light sources, with the material sample placed on a turntable which is rotated evenly to 12 different positions to achieve a denser sampling of the hemisphere of viewing directions. The cameras deliver 2048×2048 LDR images in the Bayer raw domain, which we combine to HDR images using the classical algorithm by Robertson et al. [44] and subsequently correct radiometrically. In order

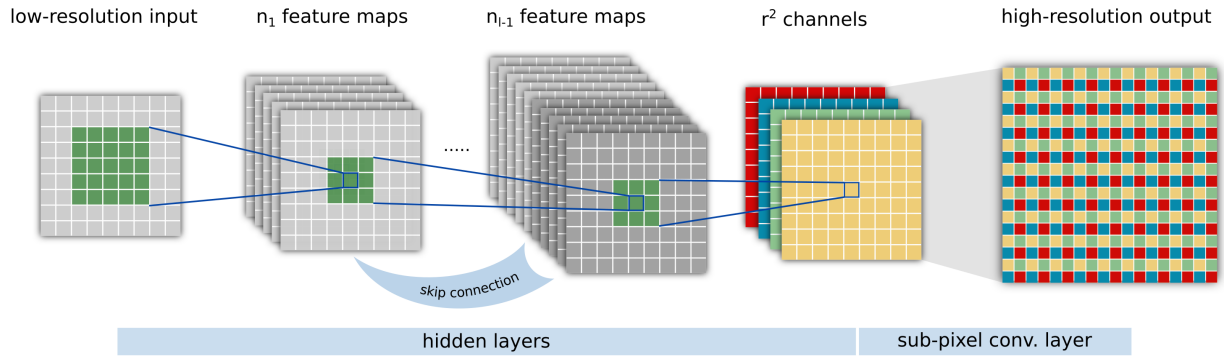


Fig. 1. The proposed network architecture is an efficient convulational neural network (CNN) with a skip connection and a sub-pixel convolution layer, extending work by Shi et al. for more detailed results on our data.

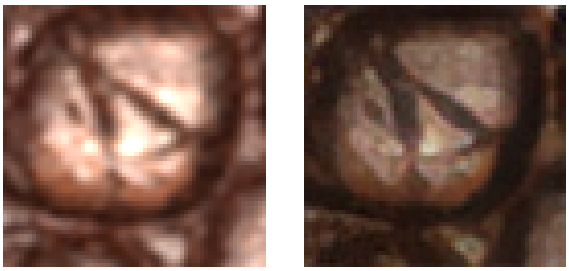


Fig. 2. Left: ground-truth. Right: reconstruction using a state-of-the-art single-image super-resolution network, created by tone-mapping the underlying HDR radiance map. Note that while the result is visually pleasing, it does not resemble the ground-truth very closely. Even if this is considered acceptable, it is unlikely that a BTF constructed this way would behave consistently under varying viewing conditions; i.e., it may exhibit flickering or similar artifacts.

to obtain full BTF measurements of the material samples in a reasonable amount of time, four material samples were measured at once, at the cost of reduced texture size, along with approximate material geometries of four 3D points per material sample (which is usually fine for BTFs). As a result, a BTF measurement for a single material consists of $n_{\text{img}} = 198 \cdot 11 \cdot 12$ HDR images of size $w_{\text{img}} \times h_{\text{img}} = 200 \times 200$. Leaving a bit of safety borders to all sides to avoid light scattering from other material samples, we end up with a texture size of $w_{\text{BTF}} \times h_{\text{BTF}} = 128 \times 128$ per material sample. Before applying the appropriate homographical transformations, the majority of the corresponding regions of interest are not axis-aligned and may therefore occupy square-shaped bounding boxes of size up to $w \times h = 200 \times 200$ in the measured HDR images. These 200×200 regions of interest are our ground-truth measurement data. We downsample these images by a factor of 0.5 to generate simulated high-resolution data, and then again by a factor of 0.5 to generate the simulated low-resolution data to be up-scaled in the training step.

5.1.2 Training/Test Split

We randomly choose two leather materials and two fabrics for testing and one leather and one fabric for validation. The remaining material measurements are used as training set.

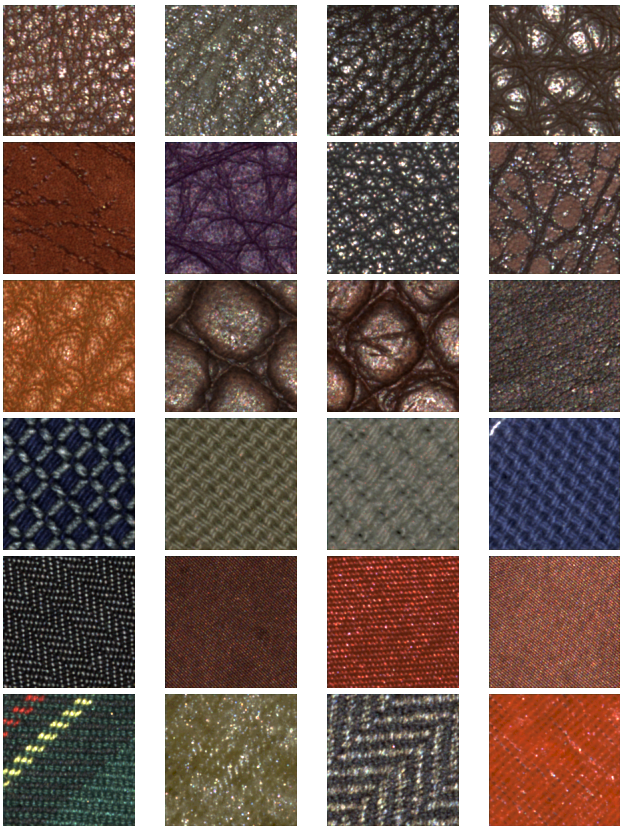


TABLE 1
Tone-mapped grund-truth HDR images from each of our database's materials.

5.1.3 Computing

Experiments are performed on a workstation with an Intel Core i7-5820K CPU at 3.30 GHz with 32 GB RAM and an NVIDIA GTX980 GPU with 4 GB of RAM, running Fedora Linux 30 and a scientific computing stack based on the Anaconda Python environment. We use PyTorch with CUDA support for our algorithm's learning step.

5.2 Implementation Details

5.2.1 Prior Art

We found that the algorithm by Dong et al. crucially depends on the presence of a non-trivial PSF. While no optical system used for BTF acquision will not meet this criterion, the cameras and lenses used for obtaining our database exhibit extremely little point spread. This is not really a problem, as in practice one can simply take the region of interest slightly out of focus. As we have no such measurements at our disposal, we just applied Gaussian blur with the parameters provided by Dong et al. [1], i.e. a 9×9 Gaussian blur kernel with $\sigma = 1.5$ to the actual ground-truth, prior to creating the down-sampled simulated high-resolution data. Our algoritm, in contrast, is applied to the original un-blurred data. For dimensionality reduction, we used rank $k = 300$ as proposed by Dong et al., which is a lot greater than what we normally use for compression ($k = 128$); we therefore expect the impact on reconstruction quality to be negligible.

5.2.2 Network Parameters

We aim for an up-sampling factor of $r = 2$. For training, we draw 17×17 non-overlapping patches from the simulated low-resolution data and the corresponding 34×34 patches from the simulated high-resolution data. In order to be able to fit the entire training data into RAM, we sub-sampled the BTFs by using approximately 5% of their bidirectional textures, which amounts to 118 images per material, or 2124 images altogether. We consider the sub-sampling justified by the large degree of redundancy in the measurements across different lighting/viewing angles, as experimentally verified by Filip et al. [45].

We implement the proposed network architecture using $L = 5$ layers, the first four of which are the convolutional layers with kernel sizes $k_1 = 5$, $k_2 = 3$, $k_3 = 3$ and $k_4 = 3$, respectively. The number of features at the respective layers are $n_1 = 64$, $n_2 = 32$, $n_3 = 32$ and $n_4 = r^2 = 4$, closely following Shi et al.

We use the Adam optimizer [46] with its default parameters and an initial learning rate of 10^-2. For the activiation functions, we choose $\phi = \mathrm{ReLU}$. Again following Shi et al., we employ the pixel-wise MSE as loss function.

5.3 Ground-Truth Comparison

For the ground-truth comparison, we use the original measured data. The proposed algorithm and the algorithm by Dong et al. are applied to the measurements down-sampled by a factor of 0.5; recall that the mapping for our method was learned on data down-sampled twice.

The learning step converged after less than 100 training epochs, which took only about an hour on the hardware used for training.

Run-time per BTF was approximately 15 minutes for Dong et al., where most of the time was spent on the re-sampling step, the computation of the SVD, and the actual algorithm.

Our algorithm has a run-time of approximately 32 minutes per BTF, 30 minutes of which are spent on the re-sampling step, which is a lot slower because it needs to be performed on the up-sampled data in our case, its run-time scaling linearly with the (linear) resolution. Using the approach described in Section 4.3, overall run-time can likely be reduced by as much as 90% depending on the choice of $\mathbf{U}$ [40], at the cost of the additional quality penalties imposed by sparse reconstruction. Given that run-time is not excessive as is, we did not investigate how much this amounts to in practice.

For comparison, we provide a number of error measures. Following Dong et al., we provide the mean RMSE

$$\mathrm{mRMSE} = \frac{1}{n_l \cdot n_v} \sum_{i=0}^{n_l \cdot n_v} \left(\frac{1}{\sqrt{w_{\mathrm{BTF}} \cdot h_{\mathrm{BTF}}}} \|\mathbf{B}_{i,:} - \widetilde{\mathbf{B}}_{i,:}\|_2 \right), \tag{8}$$

the mean PSNR

$$\mathrm{mPSNR} = \frac{1}{n_l \cdot n_v} \sum_{i=0}^{n_l \cdot n_v} 20 \log_{10} \frac{\max\left(\mathbf{B}_{i,:}\right)}{\frac{1}{\sqrt{w_{\mathrm{BTF}} \cdot h_{\mathrm{BTF}}}} \|\mathbf{B}_{i,:} - \widetilde{\mathbf{B}}_{i,:}\|_2}, \tag{9}$$

and the mean relative error

$$\mathrm{mRE} = \frac{1}{n_l \cdot n_v} \sum_{i=0}^{n_l \cdot n_v} \left(\frac{\|\mathbf{B}_{i,:} - \widetilde{\mathbf{B}}_{i,:}\|_2}{\|\mathbf{B}_{i,:}\|_2} \cdot 100\% \right), \tag{10}$$

the mean being taken over the bidirectional textures. Cf. Table 2 for the numerical errors we obtained. Numerically, our algorithm clearly outperforms the one by Dong et al., surpassing its performance in all of the metrices. Dong et al.'s results are fundamentally limited by both the compressed BTF representation they require and the smoothness prior which counter-acts the super-resolution effect to some extent.

In Table 3, we present bidirectional textures extracted from the ground-truth and the various reconstructions. Again, we obtain significantly improved results when compared to Dong et al., which lacks sharpness and exhibits dampened highlights, probably due to the compression.

Finally, cf. Table 4 for renderings of the ground-truth and the reconstructions in a standard scene for displaying BRDFs using environmental lighting.

5.4 Consistency

The renderings also provide indication for whether the BTFs produced by our algorithm are consistent in the sense that they do not exhibit noticeable artefacts caused by the per-image nature of our approach. Because of the environmental lighting, large parts of the BTF actually contribute to the overall appearance. In a scene like this, artefacts may thus become quite apparent. For the tested materials, this does not appear to be the case for either algorithm.

For further evidence, in Table 5 we provide images of ABRDFs for a single texel, which do not exhibit noticeable artifacts. In the supplemental material to this paper, we provide animated renderings of our results along with difference images, which do not exhibit artifacts like flickering, either.

5.5 Scale Invariance

Given the quality of our reconstructions, which were obtained using an up-sampling mapping learned on practically available low-resolution data, we think it is justified to conclude that, at least for an up-sampling factor of

	leather #1	leather #2	fabric #1	fabric #2
Dong	2.41e-3, 29.7 dB, 19.0 %	3.49e-3, 30.5 dB, 10.4 %	6.16e-3, 23.3 dB, 26.6 %	4.14e-3, 30.2 dB, 21.8 %
Ours	1.41e-3, 33.1 dB, 12.8 %	2.29e-3, 33.8 dB, 6.6 %	2.16e-3, 32.1 dB, 9.7 %	1.80e-3, 37.1 dB, 10.0 %

TABLE 2
Numerical reconstruction errors for the selected test materials. From left to right per table cell: mRMSE, mPSNR, mRE. Our algorithm clearly outperforms the state-of-the-art here, which relies on a lossy compressed BTF representation.

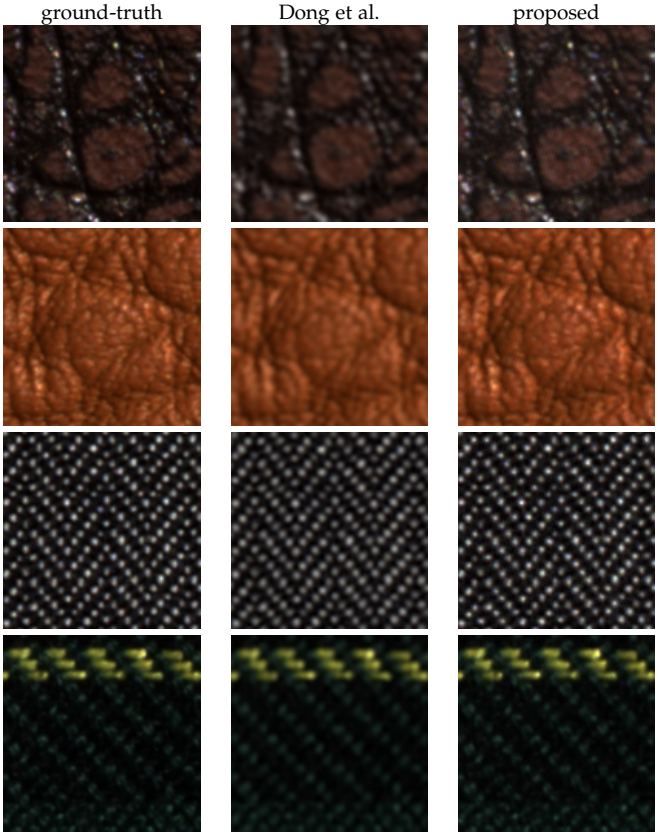


TABLE 3
Bidirectional textures extracted from the ground-truth and the reconstructions. Our reconstructions contain noticeably more detail than the state-of-the-art.

$r = 2$, our algorithm is sufficiently scale-invariant as to allow for surprisingly detailed reconstructions. For further evaluation, we conducted an additional experiment where we used a model trained to upsample data which has been down-sampled only once instead of twice to the ground-truth resolution. We observed an arguably expected slight degradation in purely numerical terms, e.g. the relative error was reduced further by 3.5% on average when compared to our actual method's model. However, the resulting images and renderings turned out to be practically indistinguishable from our previous results.

6 Conclusion

We presented a BTF super-resolution algorithm based on a simple but efficient convolutional neural network architecture. We demonstrated that, when trained on down-sampled real-world measurements, our network is capable of upsampling high-dynamic-range BTF measurements, which allows for BTFs of much higher resolution than what was measured, outperforming the state-of-the-art in BTF super-resolution. Along the way, we gave what we believe to be first, positive answers to the questions:

- Do conventional deep single-image super-resolution algorithms apply to high dynamic range data?
- Are networks trained on downsampled BTF measurements scale-invariant enough for super-resolution?
- Can we reconstruct a consistent high-resolution BTF from measurements upsampled on an image-by-image basis?

Our algorithm shares the limitations of learning-based single-image super-resolution methods. The network was chosen such that it favors faithfulness of the reconstruction over sharp, but visibly hallucinated results, which limits the achievable degree of reconstructed detail.

We think it worthwhile to investigate whether deep learning methods can be used even more effectively for our scenario. As it is, our network is not specifically tailored to BTF data. It should be possible to exploit the very specific nature of BTF measurements further, e.g. by means of a multi-view approach, or by using novel, more compact BTF representations which would result in less input channels in the network architecture. The former would be particularly useful to overcome the intrinsic limitations of single-image approaches. However, the lack of large, publically available BTF databases may prove a major obstacle on any of these paths.

Lastly, many modern super-resolution algorithms including ours are trained and/or evaluated on synthetic data generated by down-sampling in a very straight-forward manner. However, the actual image formation model is a lot more involved. Recent work by Zhang et al. demonstrates a noticeable performance improvement when taking this into consideration. [47]. We believe this to be a promising starting point for any future work on image-space super-resolution.

Acknowledgments

This work was supported by the DFG project KL 1142/11-1 (DFG Research Unit FOR 2535 Anticipating Human Behavior).

We are grateful to Clara Callenberg for help with some of the figures and Julian Iseringhausen for useful discussions.

References

[1] W. Dong, H.-L. Shen, Z.-W. Pan, and J. H. Xin, "Bidirectional texture function image super-resolution using singular value decomposition," *Applied optics*, vol. 56, no. 10, pp. 2745–2753, 2017.

[2] C. E. Duchon, "Lanczos Filtering in One and Two Dimensions." *Journal of Applied Meteorology*, vol. 18, pp. 1016–1022, Aug. 1979.

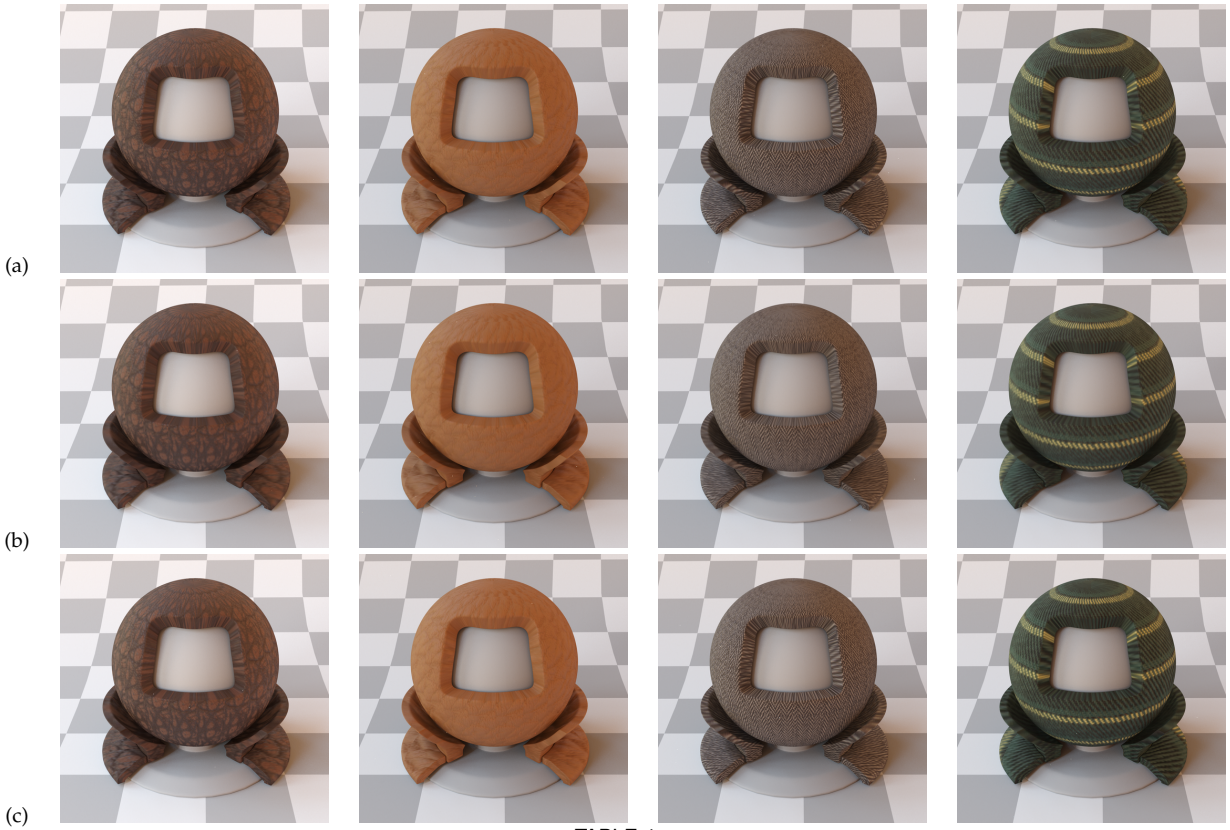


TABLE 4
Renderings of the (a) ground-truth and BTFs reconstructed using (b) Dong et al. and (c) proposed under environmental lighting.

[3] J. Sun, Z. Xu, and H.-Y. Shum, "Image super-resolution using gradient profile prior," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2008.

[4] A. Marquina and S. J. Osher, "Image super-resolution by tv-regularization and bregman iteration," *Journal of Scientific Computing*, vol. 37, no. 3, pp. 367–382, 2008.

[5] S. Dai, M. Han, W. Xu, Y. Wu, Y. Gong, and A. K. Katsaggelos, "Softcuts: a soft edge smoothness prior for color image super-resolution," *IEEE Transactions on Image Processing*, vol. 18, no. 5, pp. 969–981, 2009.

[6] Q. Yan, Y. Xu, X. Yang, and T. Q. Nguyen, "Single image superresolution based on gradient profile sharpness," *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 3187–3202, 2015.

[7] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer graphics and Applications*, no. 2, pp. 56–65, 2002.

[8] T. A. Stephenson and T. Chen, "Adaptive markov random fields for example-based super-resolution of faces," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, no. 1, p. 031062, 2006.

[9] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *International conference on curves and surfaces*. Springer, 2010, pp. 711–730.

[10] R. Timofte, V. De Smet, and L. Van Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 1920–1927.

[11] ——, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Asian conference on computer vision*. Springer, 2014, pp. 111–126.

[12] Y. Zhang, K. Gu, Y. Zhang, J. Zhang, and Q. Dai, "Image super-resolution based on dictionary learning and anchored neighborhood regression with mutual incoherence," in *2015 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2015, pp. 591–595.

[13] D. Dai, R. Timofte, and L. Van Gool, "Jointly optimized regressors for image super-resolution," in *Computer Graphics Forum*, vol. 34, no. 2. Wiley Online Library, 2015, pp. 95–104.

[14] F. Cao, M. Cai, Y. Tan, and J. Zhao, "Image super-resolution via adaptive $\ell\_\{p\}(0 < p < 1)$ regularization and sparse representation," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 7, pp. 1550–1561, 2016.

[15] J. Liu, W. Yang, X. Zhang, and Z. Guo, "Retrieval compensated group structured sparsity for image super-resolution," *IEEE Transactions on Multimedia*, vol. 19, no. 2, pp. 302–316, 2016.

[16] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 184–199.

[17] ——, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2015.2439281

[18] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, jul 2017, pp. 105–114. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/CVPR.2017.19

[19] W. Yifan, F. Perazzi, B. McWilliams, A. Sorkine-Hornung, O. Sorkine-Hornung, and C. Schroers, "A fully progressive approach to single-image super-resolution," in *CVPR Workshops*, June 2018.

[20] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1874–1883.

[21] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deep networks for image super-resolution with sparse prior," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 370–378.

[22] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in

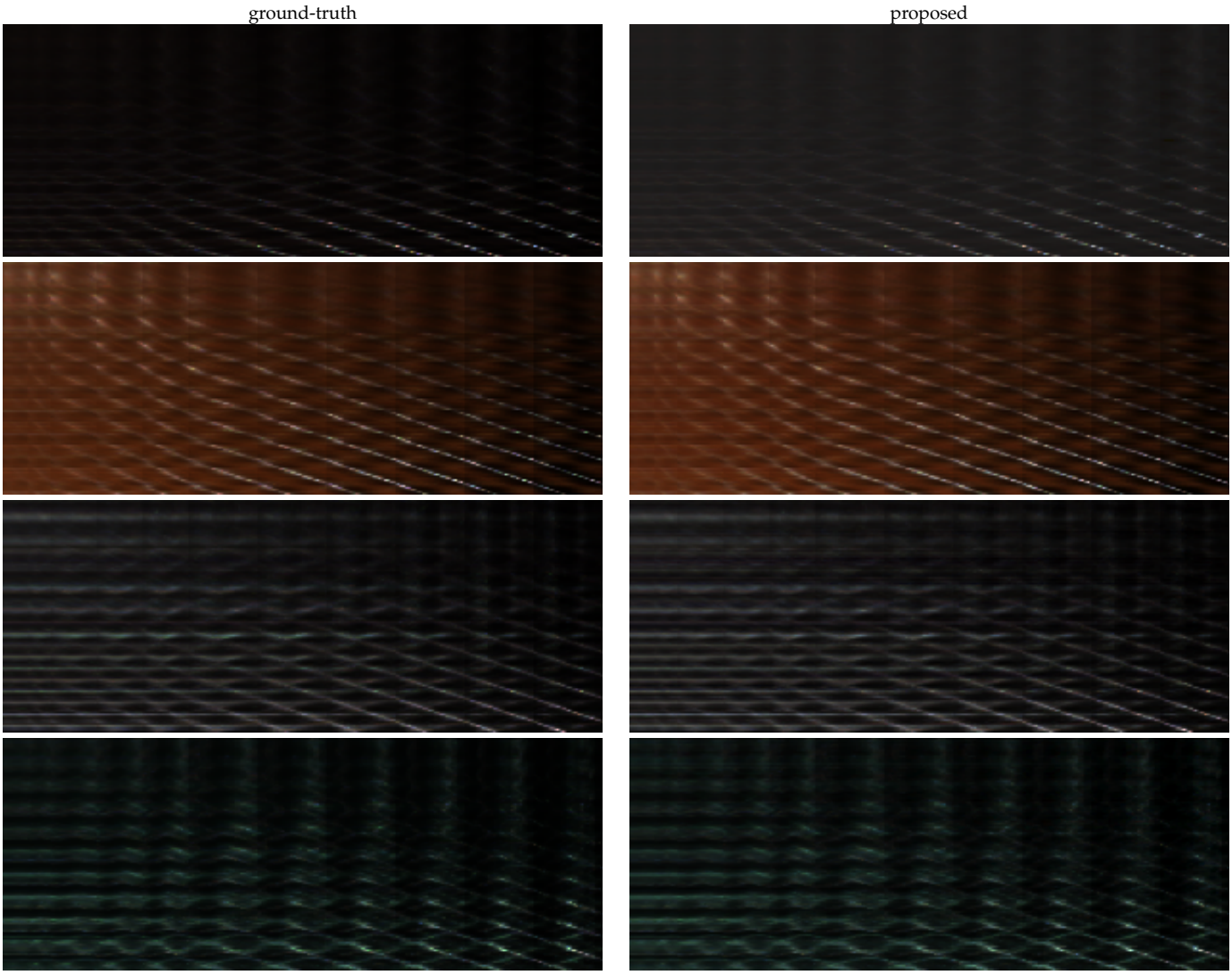


TABLE 5
ABRDFs extracted from ground-truth and our reconstructions of the four test materials. Note that the reconstructed reflectance distributions closely resemble the ground-truth, exhibiting no apparent artefacts, even though they have been obtained on a pixel-by-pixel basis.

Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 624–632.

[23] ——, "Fast and accurate image super-resolution with deep laplacian pyramid networks," *IEEE transactions on pattern analysis and machine intelligence*, 2018.

[24] Y. Wang, F. Perazzi, B. McWilliams, A. Sorkine-Hornung, O. Sorkine-Hornung, and C. Schroers, "A fully progressive approach to single-image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 864–873.

[25] M. S. Sajjadi, B. Scholkopf, and M. Hirsch, "Enhancenet: Single image super-resolution through automated texture synthesis," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 4491–4500.

[26] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C. C. Loy, Y. Qiao, and X. Tang, "ESRGAN: enhanced super-resolution generative adversarial networks," *CoRR*, vol. abs/1809.00219, 2018. [Online]. Available: http://arxiv.org/abs/1809.00219

[27] B. Wu, H. Duan, Z. Liu, and G. Sun, "SRPGAN: perceptual generative adversarial network for single image super resolution," *CoRR*, vol. abs/1712.05927, 2017. [Online]. Available: http://arxiv.org/abs/1712.05927

[28] S.-J. Park, H. Son, S. Cho, K.-S. Hong, and S. Lee, "Srfeat: Single image super-resolution with feature discrimination," in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham: Springer International Publishing, 2018, pp. 455–471.

[29] W. Yang, X. Zhang, Y. Tian, W. Wang, and J.-H. Xue, "Deep learning for single image super-resolution: A brief review," *ArXiv*, vol. abs/1808.03344, 2018.

[30] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," *CoRR*, vol. abs/1902.06068, 2019. [Online]. Available: http://arxiv.org/abs/1902.06068

[31] S. Anwar, S. Khan, and N. Barnes, "A deep journey into super-resolution: A survey," *arXiv preprint arXiv:1904.07523*, 2019.

[32] X. Zhang, J. Dong, Y. Gan, H. Yu, and L. Qi, "Btf data generation based on deep learning," *Procedia Computer Science*, vol. 147, pp. 233 – 239, 2019, 2018 International Conference on Identification, Information and Knowledge in the Internet of Things. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050919302637

[33] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, cite arxiv:1411.1784. [Online]. Available: http://arxiv.org/abs/1411.1784

[34] Y. Gan, H. Chi, Y. Gao, J. Liu, G. Zhong, and J. Dong, "Perception driven texture generation," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*, July 2017, pp. 889–894.

[35] G. Rainer, W. Jakob, A. Ghosh, and T. Weyrich, "Neural btf compression and interpolation," *Computer Graphics Forum (Proceedings of Eurographics)*, vol. 38, no. 2, Mar. 2019.

[36] K. Dana, B. Ginneken, S. Nayar, and J. Koenderink, "Reflectance and Texture of Real World Surfaces," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 1997, pp. 151–157.
[37] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct 2017, pp. 4809–4817.
[38] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR*, vol. abs/1502.03167, 2015. [Online]. Available: http://arxiv.org/abs/1502.03167
[39] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2018, pp. 2472–2481.
[40] D. den Brok, M. Weinmann, and R. Klein, "Rapid material capture through sparse and multiplexed measurements," *Computers & Graphics*, vol. 73, pp. 26–36, 2018.
[41] J. B. Nielsen, H. W. Jensen, and R. Ramamoorthi, "On optimal, minimal brdf sampling for reflectance acquisition," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 6, pp. 186:1–186:11, November 2015.
[42] W. Matusik, H. Pfister, M. Brand, and L. McMillan, "A data-driven reflectance model," *ACM Transactions on Graphics*, vol. 22, no. 3, pp. 759–769, Jul. 2003.
[43] D. den Brok, M. Weinmann, and R. Klein, "Linear models for material btfs," in *Eurographics Workshop on Material Appearance Modeling*. The Eurographics Association, 2015, pp. 15–20.
[44] M. A. Robertson, S. Borman, and R. L. Stevenson, "Dynamic range improvement through multiple exposures," in *Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348)*, vol. 3, Oct 1999, pp. 159–163 vol.3.
[45] J. Filip, M. Chantler, P. Green, and M. Haindl, "A psychophysically validated metric for bidirectional texture data reduction," *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2008)*, vol. 27, no. 5, p. 138, December 2008. [Online]. Available: http://staff.utia.cas.cz/filip/projects/pertex
[46] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," vol. abs/1412.6980, 2018. [Online]. Available: http://arxiv.org/abs/1402.6980
[47] X. Zhang, Q. Chen, R. Ng, and V. Koltun, "Zoom to learn, learn to zoom," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

Dennis den Brok studied Mathematics at the University of Bonn, Germany. After receiving his diploma in 2011, he became a member of the Computer Graphics group at Bonn University. Since 2013, he is working on his PhD, initially within the framework of the graduate school on Digital Material Appearance sponsored by X-Rite. His research interests include problems related to digital material appearance, particularly modeling, acquisition and super-resolution in several dimensions. The latter sparked his recent interest in problems from computational photography.

Sebastian Merzbach studied computer science at the University of Bonn, where he received his diploma in 2013. He is currently working on his PhD in computer graphics. His research areas are appearance acquisition, modelling and rendering, with application of deep learning methods. Other projects include spectral material acquisition and rendering, as well as material synthesis.

Michael Weinmann studied Electrical Engineering and Information Technology at the University of Karlsruhe, Germany, where he received his degree Dipl.-Ing. in 2009. After joining the Visual Computing Group at the University of Bonn in 2010, he received his PhD in computer science in 2016. His research interests include machine learning, 3D reconstruction, reflectance reconstruction, semantic scene interpretation and visualization where he published respective work on high-ranked conferences including CVPR, ICCV and ECCV as well as reputable journals such as the ISPRS Journal of Photogrammetry and Remote Sensing, ACM Transactions on Graphics, IEEE Transactions on Visualization and Computer Graphics, and Sensors.

Reinhard Klein studied Mathematics and Physics at the University of Tübingen, Germany, from where he received his MS in Mathematics (Dipl.-Math.) in 1989 and his PhD in computer science in 1995. In 1999 he received an appointment as lecturer ("Habilitation") in computer science also from the University of Tübingen, with a thesis in computer graphics. In 1999 he became Associate Professor at the University of Darmstadt, Germany and head of the research group Animation and Image Communication at the Fraunhofer Institute for Computer Graphics. Since 2000 he is professor at the University of Bonn and director of the Institute of Computer Science II.

List of Figures