Dissertation
submitted to the Combined Faculties
for the Natural Sciences and Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

put forward by
Dipl.-Inf. Burkhard Güssefeld
born in Heidelberg, Germany
Date of oral exam:

Acquisition, Modeling, and Augmentation of Reflectance for Synthetic Optical Flow Reference Data

Advisors:
Prof. Dr. rer. nat. Bernd Jähne
PD. Dr. Christoph S. Garbe

Zusammenfassung: Die vorliegende Arbeit befasst sich mit der Erfassung, Modellierung und Anreicherung von Materialreflektanzen zur hochgenauen Simulation von synthetischen Datensätzen für das Rechnersehen. Das Thema wird in drei Kapiteln behandelt: Zunächst untersuche ich die Möglichkeiten der Reflektanzerfassung, indem ich synthetisierte Bilder von hochmodernen BTF-Reflektenzfeldern mit realen Aufnahmen vergleiche. Ich zeige, dass diese als gleichwertiger Ersatz für die Bewertung von optischem Fluss verwendet werden können. Im nächstfolgenden Kapitel präsentiere ich zwei Methoden, um effizientere BRDF Modelle aus den gemessenen BTF-Daten zu erzeugen. Gemeinsam angewandt erhalten die Methoden alle relevanten Reflektanzinformationen sowie die pixelgenaue Oberflächengeometrie. Weiterhin zeige ich, dass diese Modelle zur Synthetisierung von Referenzdaten für Optischen Fluss geeignet sind und nahezu identische Ergebnisse wie die ineffizienteren BTF-Modelle liefern. Zuletzt präsentiere ich eine Methode um real aufgenommene Bilddatensätze für das Rechnersehen, mit synthetischen Niederschlagseffekten anzureichern. Dies beinhaltet eine Wasserbenetzung der Bodenflächen, Wassertropfen auf der Windschutzscheibe, sowie Spritzwasser. Verwirklicht wird dies durch eine Projektion der originalen Bilddaten auf eine rekonstruierte Szene, eine Manipulation der Szene und Oberflächenreflektanzen und eine Simulation der Lichtausbreitung der Niederschlagseffekte.

Abstract: This thesis is concerned with the acquisition, modeling, and augmentation of material reflectance to simulate high-fidelity synthetic data for computer vision tasks. The topic is covered in three chapters: I commence with exploring the upper limits of reflectance acquisition. I analyze state-of-the-art BTF reflectance field renderings and show that they can be applied to optical flow performance analysis with closely matching performance to real-world images. Next, I present two methods for fitting efficient BRDF reflectance models to measured BTF data. Both methods combined retain all relevant reflectance information as well as the surface normal details on a pixel level. I further show that the resulting synthesized images are suited for optical flow performance analysis, with a virtually identical performance for all material types. Finally, I present a novel method for augmenting real-world datasets with physically plausible precipitation effects, including ground surface wetting, water droplets on the windshield, and water spray and mists. This is achieved by projecting the real-world image data onto a reconstructed virtual scene, manipulating the scene and the surface reflectance, and performing unbiased light transport simulation of the precipitation effects.

For Mira, Ariane, and Leonie

Acknowledgements

I especially want to thank my supervisor Prof. Dr. Bernd Jähne, for giving me the opportunity for this work and always enough scientific freedom to pursue my studies. I am grateful for the stimulating and productive working environment that he created at the Heidelberg Collaboratory of Image Processing (HCI) and within the Faculty of Mathematics and Computer Sciences of the University of Heidelberg. His organization of workshops and the EMVA Bildverarbeitungs Forum always led to interesting discussions and fruitful scientific exchange.

My gratitude also goes to Dr. Christoph Garbe for agreeing to be my second advisor.

Dr. Daniel Kondermann has been a great guide throughout these years. His in-depth knowledge helped me through many challenges, and his passion for reference dataset generation steered my work and was always motivational. From him, I learned how to write papers, pursue new topics, and work in collaborations.

I am grateful for the support of Dr. Rahul Nair. He provided intelligent counsel during and after working hours and always was available for in-depth discussions on scientific and social topics.

My thanks go to the members of the HCI Ground-Truth and Light-Field groups, especially Dr. Marcel Gutsche, Dr. Katrin Honauer, Karsten Krispin, Dr. Hamza Gardi, Dr. Max Diebold, Henrik Schilling, Oliver Zendel and Alexander Brock. They have been great colleagues and table football partners. My special thanks go to Dr. Katrin Honauer, who was an excellent support for my second paper.

I also want to thank "old" members of the HCI who I, unfortunately, did not get to work with long enough, namely Dr. Moritz Becker, Dr. Henrik Schäfer, and Dr. Stephan Meister.

On a personal level, I want to thank my parents Jutta and Rainer Güssefeld for providing me all educational opportunities, teaching me curiosity, and supporting me in my scientific endeavors. Last but not least, I thank my wife, Johanna Nicolaides. Since day one, she has always been my pillar of strength, and without her incredible patience, organizational skills, and moral support, this thesis would not have been possible.

Contents

1 Introduction

1.1. General Introduction

What are the systematic differences between synthetic and real-world datasets and where do they originate? How can we improve synthetic datasets in the context of optical flow estimation? Can synthetic datasets be used to assess and increase robustness towards critical corner cases? Those are the underlying questions of my thesis. However, first, I want to review the boundary conditions of my work.

Computer vision is the research field of estimating high-level and low-level information from images or image sequences. High-level vision algorithms are concerned with extracting image semantics. This is includs object classification [14, 19, 84], object detection [125, 124, 150], and segmentation [32, 115, 134]. Low-level vision algorithms extract physical properties. For example, disparity [58, 88] and structure from motion [162, 70] algorithms can be used to estimate depth and position. Inverse rendering methods can estimate material properties or object shape [117, 91, 41], and even full virtual scene descriptions [81]. Optical Flow algorithms are concerned with estimating motion [100, 133].

As for any estimation methods, accurate reference data is needed in the development and validation process of computer vision algorithms. With the recent emergence of Deep Neural Networks for most computer vision tasks, the need for reference data is further amplified for training and testing purposes.

Creating accurate computer vision reference datasets is a costly, challenging, and a time-consuming task. Manual labeling is typically used to generate ground truth for high-level semantics or object detection [21, 31]. Labeling low-level information on a pixel scale can be achieved, but is complex and severely limited by the accuracy of the manual process [25]. Expensive measurement devices, *e.g.*, laser scanners and state-of-the-art camera rigs, are used to create large-scale reference datasets like

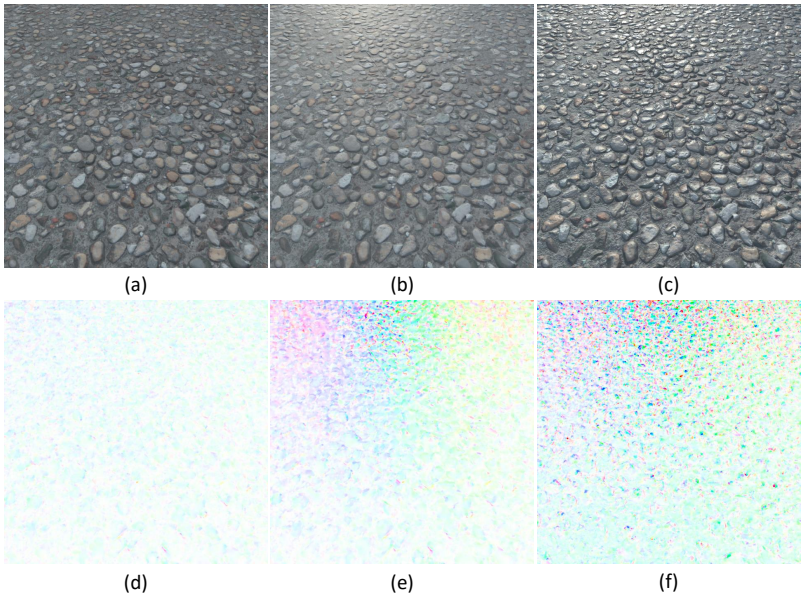

Figure 1.1.: **Toprow:** Synthetic renderings with varying degree of realism, a skylight with the sun from the front, and a scanned physically based gravel ground surface from Substance Source[1]. **Bottomrow:** The corresponding optical flow computed for a still image-pair where the light source was rotated clockwise by 10 degrees. HSV color-coding of Middlebury [6] is used, where white depicts a correctly estimated flow of zero. The image in (a) visualizes a diffuse rendering that conforms with the optical flow brightness constancy assumption. Consequently, the flow errors in (d) are minimal. Manually adding a spatially constant specular term in (b) incurs additional errors to the flow fields (e). However, the realistic appearance of the scanned material (c) significantly increased specular and geometric details and causes flow errors in (f) with a vastly different, higher-frequency spatial distribution. Most datasets use models as in (a) and (b). Consequently, the confidence of the datasets is decreased and systematical differences to real-world images occur.

KITTI [38] or the HCI Benchmark Suite [71]. However, the accuracy of such ground truth is inherently limited by measurement errors [72]. For the algorithm class of optical flow, the motion ground truth is further restricted to that of static or rigidly moving objects within engineered scenes [6]. By contrast, synthetic datasets, generated with computer graphics, have perfect ground truth available. Additionally, computer graphics facilitate the ability to parametrize the scene conditions, such as weather and lighting, at will, which in turn allows creating a large number of varying image sequences.

Synthetic datasets are particularly enticing for the development process of flow algorithms, due to the lack of accurate real-world ground truth. However, systematic differences were identified between the performance on real-world and synthetic datasets, due to lighting, scene composition, and the general appearance of the images [147, 101]. Throughout the work presented in this thesis, I address this problem. However, first, I briefly explain where the problems originate and how I approached a solution.

The underlying assumption of optical flow methods is that of brightness constancy, *i.e.*, a diffuse world with static lighting. Consequently, scene conditions that severely violate brightness constancy are critical optical flow corner cases, *i.e.*, reflections, the motion of high-frequency shadowing effects, and light transmission. As depicted in Figure 1.1, the performance of optical flow algorithms highly depends on the spatiotemporal dis-

tributions of image intensities. A realistic and discriminating synthetic optical flow dataset should, therefore, simulate all dynamic appearances changes accurately.

Ignoring all camera aspects, a topic which I do not address in this thesis, appearance is the product of two aspects (*c.f.*, Figure 1.2): First, the light sources of the scene. Second, the reflectance properties of scene materials. While the former mostly exhibits low-frequency changes over space and time, the latter can convolve the incident light to high-frequency spatiotemporal appearance. A light transport algorithm simulates this convolution. For creating realistic and discriminating synthetic reference data, which best conforms with reality in spatial and temporal domains, a physically based light transport algorithm [118] is, therefore, an essential requirement and physically accurate reflectance data is critical.

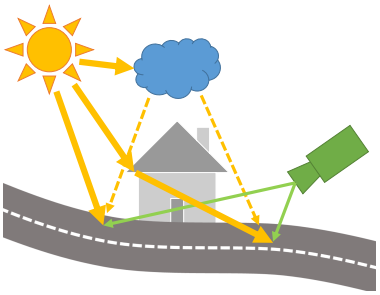

Figure 1.2.: Simulating the appearance of a scene is a joint problem. The appearance of an observed surface depends on all incident light and its reflectance properties. Light can be directly incident from light sources, or via reflection or scattering of other surfaces and volumes. Typically, light is emitted with low-frequency changes over space and time, *e.g.*, consider the slow changes of the sunlight. However, light-matter interaction can create higher-frequency spatiotemporal appearance changes, *e.g.*, from specular highlights or the shadows of dynamic objects. Physically based light transport can simulate even most complex light paths [118]. However, physically plausible descriptions of surface reflectance are required to create critical high-frequency appearance effects with high realism. In this work, I explore methods for accurately recovering surface reflectance from measurements and the impact of various reflectance effects on optical flow performance.

However, existing synthetic datasets, such as [26, 16], predominantly use physically implausible rendering systems without separation of lighting and reflectance. More precisely, lighting information is statically precomputed and encoded in the material textures. Additionally, the utilized textures are artistically driven and do not represent real-world reflectance. As a result, the simulated are unrealistic and dynamic effects from the real-world become static and thus unchallenging to optical flow algorithms.

In the following chapters, I explore ways of accurately capturing and modeling the material reflectance to improve the confidence of synthetic reference data for optical flow algorithms. I further show that by using physically based material models, it is possible to augment the appearance of materials in a plausible way. I conducted all my experiments with unbiased path tracing to ensure that no error is introduced from the light transport algorithm.

1.1.1. Topics

The topics of this thesis are *Reflectance Field Renderings For Synthetic Optical Flow Datasets*, *Acquisition and Benchmarking of BRDF Models For Synthetic Optical Flow Datasets* and *Augmentation of Physically Plausible Precipitation Effects*. They relate to the problem stated above and are summarized in the following paragraphs:

Reflectance Field Renderings For Synthetic Optical Flow Datasets The motivation for the work of this chapter arose from the observation of insufficient realism in existing synthetic datasets. In this context, I explored ways of accurately capturing surface reflectance. I show that by using state-of-the-art reflectance field measurements, the quality of synthetic datasets can be sufficiently improved. Compared to datasets captured with a real-world camera, the synthesized datasets are almost indistinguishable to optical flow algorithms. The remaining minor differences can be accounted for by increasing the sampling density of the capturing process.

Creating Compact BRDFs From Reflectance Fields For Synthetic Optical Flow Datasets Reflectance Field renderings are computationally expensive and infeasible for real-time applications. Analytical BRDF models can, however, be used to model physically accurate reflectance properties. I show that by fitting BRDF models to the reflectance fields, the quality of synthetic datasets can be maintained. I further identify and thoroughly assess the most critical reflectance effects for optical flow performance analysis and propose the minimal BRDF quality for various material types. The resulting BRDF textures can be applied to real-time and offline rendering systems alike and are efficiently rendered.

Augmenting Real-World Datasets with Physically Based Precipitation Effects Not for all computer vision tasks, purely synthetic datasets are always desired. On the other hand, real-world datasets are limited with respect to variety and often lack critical corner cases. I present a method for augmenting real-world datasets with ground surfaces wetting, water droplets on the windshield, and water particles in the form of mists and water spray. In contrast to existing phenomenological approaches, the proposed method augments the original image data in 3D world space and is, therefore, capable of producing physically plausible results.

2
Background

In this chapter I present the conceptual and theoretical background for this thesis. Dense optical flow estimation methods are the computer vision algorithms at issue. In Section 2.1, I present their underlying principles and how those lead to a dependency on statistical brightness variations in the images. Synthetic datasets are a commonly used tool in the development and validation process of optical flow algorithms. Synthesizing images with physically realistic brightness variations can increase the confidence of the datasets.

The required methods for simulating physically realistic datasets fall within the scope of computer graphic domains. The virtual scene creation, outlined in Section 2.2, is concerned with the creation of virtual geometry, animations, and material textures. The rendering process simulates the propagation of light and is presented in Section 2.3. Material models are the link between the two domains and the central topic of this thesis. Section 2.4 details the material models of opaque surfaces and the process of capturing the material textures. In Section 2.5, I present the theoretical background on optical properties of wet materials and an efficient, physically based wetness model.

The structure and depth of the chapter were chosen to give an overall picture and better understanding of simulations in the context of synthetic optical flow datasets. For more detail, I refer to standard literature.

2.1. Optical Flow

Optical flow describes a class of computer vision algorithms concerned with the dense, *i.e.*, pixel-wise, estimation of motion in images of an image sequence. In the strict, original usage of the term, optical flow is understood to be the apparent motion of pixels between subsequent images. Today, however, the term is used more frequently as a synonym for *Motion Estimation*,

which is a class of methods that estimate the *actual* motion of pixels [133]. While the apparent and actual motions of objects are the same for objects with Lambertian surfaces, they can be very different for reflective surfaces where the apparent movement is that of the reflections whereas the actual movement is that of the reflective object itself.

The applications of optical flow and motion estimation are plentyful; From 3D reconstructions using structure from motion techniques [8] over object tracking [83] to video compression[78] and frame interpolation [158]. For video compression and frame interpolation, the estimation of apparent motion is desired as the goal is to fool human perception. For 3D reconstruction and tracking, in most cases, the actual motion is the one required as this actual motion can be used together with triangulation techniques to reason about the 3D world.

The class of methods that my work impact the most are those in Photogrammetry and 3D reconstruction. Therefore, I follow the convention of using optical flow as a synonym for 2D Motion estimation in this work.

The majority of model-based methods to compute optical flow have a brightness constancy assumption. That is:

$$I(x, y, t) = I(x + u(x, y), y + v(x, y), t + \Delta t) \tag{2.1}$$

The intensity/color I of an pixel at (x, y) that moves to $(x + u(x, y), y + v(x, y))$ between time t and $t + \Delta t$ ts is the same. This leads to an optimization problem of finding $u(x, y)$ and $v(x, y)$ such, that $|I(x, y, t) - I(x + u(x, y), y + v(x, y), t + \Delta t)|$ Solving this for every pixel independently leads to the *aperture problem* of optical flow. Consider a moving white wall observed through a small aperture. We can not tell in which direction the wall is moving. Given a broader context (perhaps the corners of the wall) we can suddenly discern the movements. To circumvent this problem, two classes of optical flow methods have evolved that utilize brightness constancy in some way or the other: **Feature based methods** estimated sparse optical flow for image features, *i.e.*, salient points where the image neighborhood is easily distinguishable. Here image features are extracted and tracked between two consecutive frames. An affine transformation model is then fitted to the feature tracks. Popular feature-based methods are the Lucas-Kanade method [86], which estimates optical flow of feature points and its local neigh-

borhoods, and the Buxton and Buxton method [17], estimating the optical flow of edges. While feature-based methods are fast and most robust to image noise [7], they only estimate sparse flow fields and do not work for uniform image regions.

Pixel based methods estimate dense optical flow fields, defined for each pixel of an image. Instead of using a rigid image neighborhood around a feature point that is correlated with the second image, these methods minimize a term consisting of the original brightness constancy term above and an additional regularizer, that ensures that neighboring pixels have similar values:

$$u, v = \underset{u,v}{\operatorname{argmin}} \int_{x,y} \Phi(I(x,y,t) - I(x + u(x,y), y + v(x,y), t + \Delta t)) + \gamma R(u,v) \quad (2.2)$$

where the Φ is the data term, usually a L1 or L2 norm, γ is a scaling factor and R is a regularizer that ensures a smooth flow field. For example:

$$R(u,v) = \int_{x,y} \nabla u(x,y) + \int_{x,y} \nabla v(x,y) \quad (2.3)$$

Different choice of data term and regularizer lead to different objective functions with different optimization strategies [62][166]. I refer to [6] for further information on dense methods.

A final class of methods exists, that is increasingly popular, consists of learning based methods. If the actual motion is of interest, brightness constancy can be violated for all non-Lambertian surfaces, and a plethora of research is done in making optical flow methods robust towards violations of this assumption. Learning approaches [157], [64] aim to overcome this issue by learning the actual motion model. While the specifics of each learned model are different, they all have in common that a large dataset of reference data consisting of image pairs and reference motion fields is used as training data to a learning algorithm. The trained model can then predict motions fields on previously unseen image pairs. Apart from designing the learning system itself, the biggest challenge is to obtain the amount of reference data required for training to succeed. Synthetic data is one viable option for generating such large amounts of data required [142], [16] Like any other dataset, it needs to be ensured that the synthetic data covers all aspect of the application do-

main, where the flow method is to be utilized, *e.g.*, specularities, reflections.

2.2. Creating Virtual Scenes

This section describes the basic computer graphics techniques, that can be used to create virtual scenes. Such virtual scenes can, in turn, can be used by rendering systems to synthesize images. The **Modeling** and **Animation** processes define the shapes and temporal descriptions of a scene. The **Texturing** process is concerned with providing inputs for the material models of the target rendering systems.

2.2.1. Modeling

Modeling is the computer graphics discipline, concerned with creating descriptions of object shapes [33]. These descriptions are also called geometric models or 3D-models. They can consist of curved surfaces, voxels, or vertices that are connected to polygonal meshes. Polygonal meshes are the most prevalent geometry representation, as GPUs can efficiently process them. They can be created in different ways.

Artistic modeling is the process of assembling primitives, *e.g.*, cubes or planes, to complex three-dimensional shapes. A 3D modeling software tool, such as Blender [11], is used in this process (cf. Figure).

Procedural modeling can be used to create three-dimensional shapes from a set of rules [28]. Typical rule sets use L-Systems [127], generative modeling [80] and fractral noise. It is best suited for structured or organic shapes, terrain, or buildings.

Scan based modeling Laser scanners, structured light scanners or cameras in conjunction with structure from motion [162, 70] methods, can be used to scan the geometry of real-world objects. These methods produce point clouds, that can be transformed into polygonal meshes [69, 140]. However, reconstructed meshes often exhibit artifacts. Scanning noise can create disruptive patterns on the surface. More importantly, concave objects posses ambiguities, when local neighborhoods of points are estimated. As a result, scanned meshes usually have to be cleaned up after reconstruction.

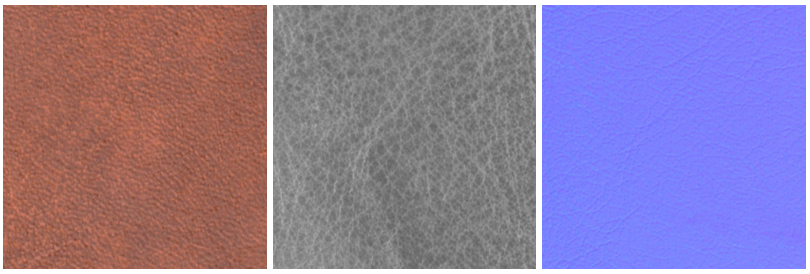

Figure 2.1.: Exemplary leather input textures for the physically based material model of Disney [15]. The diffuse color or albedo (a) defines the view-independent diffuse appearance of the model in sRGB colors. The roughness texture (b) defines the distribution of the view-dependent specular term, where black corresponds to mirror-like reflectivity and white corresponds to a diffused reflectivity. The normal map (c) defines the mesoscopic surface geometry on the pixel scale. A light blue (0.5, 0.5, 1.0) corresponds to an upright surface normal, a red color of (1.0, 0.5, 0.5) marks a normal pointing along the x-direction, and a green color of (0.0, 1.0, 0.5) marks a normal pointing along the y-direction.

2.2.2. Texturing

Texturing is the process of creating inputs for material models, that can be applied to geometric models via UV-mapping [33]. A UV-map defines the location of each vertex on the texture set. The UV-map may be defined by hand or automatically by projection. Typically, texture sets at least include a texture for the diffuse term and one or more textures for the specular term. An example texture subset for a leather sample and the physically based material model of Disney [15] is depicted in Figure2.1. It should be noted, that while a BRDF model does only describe the microscopic properties of the surface, often a normal map is included to model the *mesoscopic geometric detail* on a pixel scale. This mesoscopic geometric detail increases the perceived geometric detail of the otherwise coarse 3D geometry.

Artistic texturing is the process of manually painting the texture inputs. To this end, texturing tools can be used to paint the textures on the meshes directly. The process is error-prone, as physically implausible texture sets can quickly violate physical principles.

Procedural texturing is the process of sampling textures from mathematical descriptions. This can include but is not limited to sampling fractal noise, cellular textures, and predefined texture patterns [28]. Defining the mathematical description is, in most cases, more time consuming than artistic texturing. However, it is simple to create variations of texture sets, and physically plausibility can be ensured more easily.

Scan based texturing is the process of creating textures from measurements. Photometric stereo can estimate the surface normals and diffuse Lambertian reflectance from images taken under varying lighting conditions [90]. More sophisticated approaches use multiple lighting as well as viewing directions to estimate full sets of material textures. For a detailed assess-

ment of accurate measurement techniques and devices, I refer to section 2.4. Recently, methods have been proposed, that approximate a full texture set from a small set of images by means of deep learning [24, 2, 82]. It should be noted, that even though they are capable of producing plausible texture sets, they can severely diverge from accurate measurements for arbitrary viewing and lighting directions.

2.2.3. Animation

Animation is not a core issue of the work presented in this thesis. I will, therefore, only briefly cover the concepts used throughout my work. For more detail, I refer to [33].

Animation is the process of defining motion and temporal descriptions of scene objects. The objects may include three-dimensional shapes, cameras, lights, and textures. In most cases, keyframing is used to define the object or texture state at discrete timestamps. Intermediate timestamps are then interpolated linearly or with splines. The keyframed states can be defined via artistic animation or by capturing and tracking the motion of real-world objects [164, 165] and cameras [72].

2.3. Rendering

Rendering is the process of synthesizing a discrete, digital raster image for a virtual sensor placed in a virtual scene. The output of the rendering process is also called render or rendered image.

Rendering has many applications. This includes visual effects, video games, design and visualization, and simulations. Depending on the application, the rendering process has to meet different requirements. Naturally, for real-time applications, *e.g.*, video games, the rendering speed is the decisive factor. The rendering quality can only be increased to the extent that current consumer GPUs are still able to ensure 30Hz or higher framerates. To this end, real-time rendering systems use harsh approximations, such as rasterization and precomputed lighting. Other rendering applications, *e.g.*, design and visualization or visual effects, have lower time constraints. Here, offline rendering systems that allow the accurate simulation of light-matter interaction are most prevalent. These type of physically accurate rendering systems are also useful for optical flow simulations to ensure that no rendering bias affects the simulation quality.

The rendering equation is the theoretical background of creating physically accurate renderings. It formulates the process of rendering of a digital raster image as a light transport problem. For a more detailed description, see section 2.3.1.

A rendering system computationally solves the light transport problem. It consists of four major components that correspond to components of the rendering equation. The three components are:

1. The **light transport algorithm** which integrates the collection of light paths.
2. The **lighting model** which models how light is emitted from light sources.
3. The **BSDF model** which models how light interacts with surfaces and volumes.
4. The **camera model** which models how light interacts with the optical system and the sensor.

The quality of the **light transport algorithm** is determined by the bias, or error, with which the light transport integral is approximated. While there are no analytical solutions for the integral, Monte-Carlo methods can sample the integral without bias. Unbiased and biased light transport algorithms shall be discussed in more detail in 2.3.2.

The quality of the **lighting model** is determined by the accuracy with which radiation characteristic of real-world light sources can be described. This can include using radiometric light units, accurate directionality of light, *e.g.*, from IES profiles, and accurate daylight systems, *e.g.*, the Hosek *et al.* sun and sky model [63]. While physically realistic lighting models are certainly an interesting topic and linked to general simulation quality, they fall out of the scope of this work and shall not be discussed in more detail.

The quality of the **BSDF model** is determined by the ability to describe physically plausible light-matter interactions. According to the geometric optics abstraction, BSDF models should cover absorption, reflection, scattering, and transmission of light within a surface or volume. However, BSDF models can also support some wave effects of light, such as the polarization of light through Fresnel reflection and iridescence [10]. This shall be discussed in more detail in 2.3.3. For opaque materials, the

BSDF can be trimmed by its transmission term and simplified to the BRDF. Since reflectance modeling of opaque materials is at the core of this work, BRDF models and their mathematical background are discussed more thoroughly in section 2.4.

The quality of the **camera model** is determined by the ability to match the optical imaging properties of real-world optics and sensors. Real world cameras boast various aberrations which are caused by optical systems and sensors. The optical aberrations manifest themselves as image blur and image distortion. The blur can be modeled by a spectral Point Spread Function (PSF), describing how the spectral radiance of a virtual point light source is spread over the virtual sensor plane. The optical distortion can be modeled with a polynomial distortion model or a distortion map. Both models describe how the exit direction of camera rays diverge from a perfect pinhole camera. Sensor aberrations include sensor noise and shutter artifacts, such as motion blur and the rolling shutter effect. In contrast to sensor noise, which can be efficiently applied in a postprocessing step, shutter effects have to be handled during the rendering process due to the dependency on spatiotemporal events. The applicable camera models in computer graphics are the pinhole and thin-lens model. Compound lenses are technically possible but, to my knowledge, rarely used. Real-time rendering systems can only render pinhole cameras inherently due to technical constraints of rasterization. Camera specific effects can only be approximated and applied in a post-processing step. On the contrary, physical light transport simulation with a thin lens camera model allows the simulation of some camera effects during the rendering process. More precisely, depth-of-field and distortion caused by the lens setup and aperture of the optical system and the shutter effects from the sensor. However, the wave related portion of the PSF is missing, due to the lack of interference simulation in the geometric optics light transport. Just like the lighting model, computer graphic camera models are not directly linked to the core topic of this work and not discussed in more detail.

In the past, rendering systems have often been developed without separation of lighting and materials, due to computational constraints, lack of knowledge, or to provide artistic freedom. The drawbacks of such rendering systems are that textures and lighting do not have physical meaning, have to be handcrafted to match the rendering system, and cannot be transferred to other

systems. While such empirical rendering system can create the human perception of photorealism, it is not possible to create physically realistic simulations for arbitrary lighting and scene setups.

Physically based rendering (PBR) is a concept for building rendering systems and aims to remove this deficiency [118]. It combines physically accurate light transport algorithms, material models, and lighting models with measurements and physical rules that guide the creation of textures and lighting of virtual scenes. Physically based rendering has been a recent trend for offline rendering as well as real-time engines and is the basis for my work on augmenting physically based wet surfaces. For details about this concept, see section 2.3.4.

2.3.1. Light transport

Like any electromagnetic wave, light and its behavior are studied in the physical domain of electromagnetism. At its core, light generation and interaction can be described by the Maxwell Equations. More precisely, the Maxwell equations model the interaction between electromagnetic waves as well as the interaction of electromagnetic waves with conductive and dielectric materials. While it is possible to solve the Maxwell Equations computationally, it is computationally expensive and only feasible for small scale problems.

The geometric optics approximation models light transport as a propagation of rays [44]. It is accurate when the observed structures are magnitudes larger than the wavelength of light. Ray propagation is computationally manageable, even for large and complex scenes. Therefore, the geometric optics approximation is the best practice for simulating light-matter interactions in computer graphics. When needed, the geometric optics approximation can even be extended to partially support some wave optics properties, such as iridescence [10] and polarization [159].

Each pixel in a rendering depicts the appearance of volumes and surfaces observed through the aperture angle of the sensor pixel. In radiometric terms, this corresponds to the total amount of irradiance, measured in $\frac{W}{m^2}$, that is incident per solid angle of the sensor. This is equivalent to the average amount of radiance, measured in $\frac{W}{m^2 sr}$, that all 3D points in space seen through the pixel reflect or transmit towards it. The rendering equation [67]

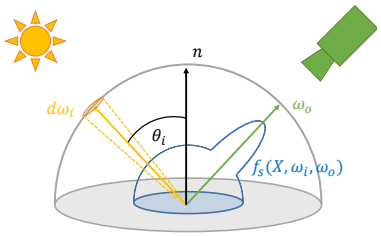

Figure 2.2.: The Rendering Equation formulates rendering as a light transport problem. Light is transported from light sources, past materials, and towards a viewer. The incident irradiance on a surface point X depends on the radiant flux, the observed solid angle $d\omega_i$ from X and the elevation angle Θ_i of the light source. A material-specific BSDF function models the interaction of light with the surface point. It relates the outgoing radiance in viewing direction ω_o to the incident irradiance.

describes the radiance $L(X, \omega_o)$ of a 3D point in space X, which is reflected or transmitted towards an observing direction:

$$L(X, \omega_o) = L_e(X, \omega_o) + \int_\Omega f_r(X, \omega_i, \omega_o) L(X, \omega_i) \cos\Theta_i d\omega_i \quad (2.4)$$

where

- ω_i, ω_o are the incoming and outgoing light directions
- Ω is the unit hemisphere around the point X
- $L_e(X, \omega_o)$ is the emitted radiance towards ω_o
- $L_($X, \omega_i$) is the radiance received by the point X from ω_i
- $\cos\Theta_i$ is the factor of the visible surface area from ω_i
- $f_s(X, \omega_o, \omega_i)$ is the bidirectional scattering distribution function (BSDF) that relates the outgoing radiance to the incident irradiance. It describes what ratio of incoming light is reflected or transmitted towards the observer ω_o

In practice, the rendering equation usually models spectral radiance by adding a dependency on the wavelength. For clarity, this dependency has been omitted.

The rendering equation formulates the problem of rendering a digital raster image as a light transport problem consisting of many light paths. A visual example of such a light transport path for an opaque material is depicted in 2.2. A light source emits radiant flux into its environment. An opaque surface point receives a certain amount of irradiance from the light source, proportional to the solid angle and the visibility factor of the surface. The solid angle depends on the size of the light source and its distance. Less light is received from the surface when it is not parallel to the light ray, and the unit projected-area is decreased. The visibility factor models this phenomenon, and depends on the unit projected-area of the light and thus the elevation of the light source. The light is then absorbed, scattered, and reflected by the surface material. The fraction of light reflected towards the viewer is determined by the material specific BSDF and the solid angle of the viewer.

Light transport algorithms can be used to approximate or sample the integral for a given scene. Several available algorithms and their impact on the simulation accuracy are discussed in the next section 2.3.2.

2.3.2. Light Transport Algorithms

Light transport algorithms are the central algorithms for synthesizing images with a rendering system. They are responsible for simulating the received radiance for each pixel of a virtual camera placed in a virtual scene. Thereby, they take into account the composition of light sources, geometry, materials as well the virtual camera properties. The rendering equation 2.4 formulates the rendering process as a light transport problem. Its light transport integral cannot be solved analytically since the incident irradiance of each point in space depends on an infinite amount of light paths. Therefore, light transport algorithms either solve a simplified light transport problem or estimate the light transport integral by sampling light paths.

Rasterization (with programmable shading) projects the vertices of the scene geometry into the raster of a virtual pinhole camera. The depth of the vertices is used to perform a visibility check on the geometry. The pixels of the image are then shaded based on the light sources of the scene and the positions, material textures, and shading normals of the underlying projected vertices. Vertices do not possess information about other vertices in the scene. Therefore, global light transport, such as reflections or indirect light, cannot be computed. Rasterization with programmable shading essentially solves a simplified light transport problem, that only considers direct diffuse light paths. It can be efficiently implemented and parallelized on GPUs and is, therefore, the standard real-time light transport algorithm. For the sake of completeness, it should be noted that modern real-time rendering systems posses advanced shading techniques to approximate global light paths in screen-space. However, more than a single bounce of light cannot be approximated with screen-space techniques.

Ray Tracing is a simulation tool to compute the path for waves and particles. Typical applications in other research fields are geometric optics simulations, such as optical design or, radio signal simulations and acoustical simulations. Ray tracing can also be used to render images. Rays are shot through pixels of the virtual camera and then propagated, or traced, through the virtual scene. When rays hit a surface of an object, a shading event is triggered: The BSDF of the surface is evaluated, and new rays

are spawned towards light sources or into the scene. Fundamentally, ray tracing is a method to sample light paths of the light transport integral in reverse. The backward traversal is valid, due to the Helmholtz Reciprocity of light propagation. When trying to estimate the full light transport problem accurately, ray tracing can quickly become computationally expensive. Theoretically, an unlimited amount of light rays would have to be spawned at every surface interaction, each being traced with an unlimited tracing depth. In practice, this problem is solved by limiting the tracing depth, the amount of spawned rays per surface interaction, and the possible light paths. Ray tracing approaches that can sample all light paths of the light transport problem use Monte-Carlo sampling.

Monte-Carlo Ray Tracing Monte-Carlo sampling can be applied to ray tracing to limit the number of active light paths when estimating the full light transport problem. The basic Monte-Carlo method for sampling light paths is **path tracing** [67]. It utilizes Monte-Carlo sampling for two stages of light-matter interactions:

1. Sampling the BSDF: A random BSDF event is chosen at every interaction. The choices include absorption, reflection, transmission, and scattering. An absorption even terminates the light path. The reflection, transmission, and scattering events spawn new rays along with their corresponding directions.

2. Sampling the ray direction: A random ray direction is sampled along the hemisphere of the surface or sphere of the volume. Importance sampling can be used to ensure that the ray directions are sampled according to the distributions of the BSDF components.

Path tracing is theoretically capable of simulation all light paths. It can, therefore, estimate the full light transport problem. However, the probability of sampling light paths with a significant contribution to the light transport integral can be low. This creates noisy renderings for complex scenes with small light sources or highly specular materials. Several methods can be used to combat rendering noise. They include next event estimation [135], bidirectional path tracing [149], photon mapping [66], metropolis light transport [148] and path guiding [102, 151].

Render Bias The render bias is the systematic error of the estimated solution of the light transport problem. A light transport algorithm is unbiased if for any sample count the estimated solution of the light transport problem is without render bias. In other words, when averaging an infinite amount of estimated renderings of the same virtual scene, an unbiased light transport algorithm produces the correct result. Biased light transport algorithms are consistent when the render bias converges against zero in the limit of the rendering process. In other words, when rendering an image of a virtual scene for an infinite duration, a consistent light transport algorithm produces the correct result. Unbiased light transport algorithms are always physically accurate on a geometric optics level. Consistent algorithms can be physically accurate, but require an error measure to ensure that the bias is sufficiently low.

Even though **rasterization** with programmable shading does not technically sample the light transport integral, it can be considered biased, as the rejection of light paths introduces a systematic error in the renderings. It should be noted, that recent advances in GPU design allow ray tracing specific light paths during the rendering [50]. This reduces the bias of the rendering process, but cannot reduce it to zero.

In practice, **ray tracing** introduces render bias by limiting the length of light paths, the amount of spawned rays per light-matter interaction, and by light path rejection. It is also inconsistent, as infinite sampling cannot guarantee that all light paths are covered.

Path tracing and most of its variants are unbiased methods since all light paths can be sampled. **Photon mapping** introduces bias by gathering photons in a finite radius. However, it can be made consistent by extending it to stochastic progressive photon mapping [48], which can ultimately be combined with bidirectional path tracing [40] [73]. The remaining render bias can then be assessed by the gathering radius of the final sampling.

2.3.3. Light Matter Interaction

Light, that is not propagated in a vacuum, interacts with its surrounding matter. Light can propagate within dielectric materials at the speed of light, that is inverse proportional to the index of refraction (IOR) of the dielectric material. When encounter-

ing particles of the dielectric material, light either gets absorbed or scattered by the particle. Conductive materials are usually dense enough, so that light cannot travel considerable distances within the material. The energy of the light is absorbed and reemitted by the particles in a material specific spectrum.

When light hits an interface between two different media, two things can happen:

1. The light is **reflected** along the perfect mirror direction into the backward half space. Light reflection, on a geometric optics level, occurs for every physical material. The portion of reflected light depends on the IOR difference of the material interface and the incident angle of the light. It increases towards grazing angles and can be computed with the Fresnel Equations. Dielectrics usually reflect only a small portion of the light at normal incidence. The remaining portion is refracted inside the material. Metals reflect most of the light at normal incidence and absorb the remaining portion.

2. The light gets **refracted** into the forward half space and the direction is changed, due to a modulated speed of light. Snell's law describes this behavior. Refraction only happens for dielectric materials.

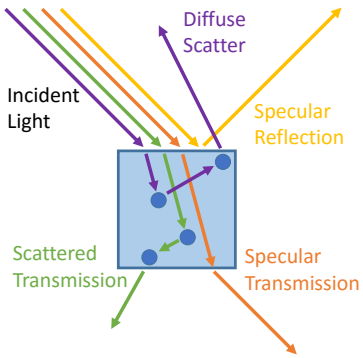

Figure 2.3.: The BSDF is a scattering function that can be used to describe light matter interaction on a geometric optics level. It consists of a scattering reflection term (BRDF) and a scattering transmission term (BTDF). The BRDF models reflection paths (yellow) and scattering paths into the backward half space (purple). The BTDF models refracted transmission paths (orange) and scattering paths into the forward half space (green).

All light-matter interaction is modeled by the Bidirectional Scattering Distribution Function (**BSDF**). It is a material-specific function, that describes light reflection, absorption, and scattering properties on a geometric optics level, see figure 2.3. It covers translucent dielectric materials, such as water, mostly opaque materials, such as wood or stone, and conductive materials, such as metals. The BSDF can be split into two components. The Bidirectional Reflectance Distribution Function (**BRDF**) and the Bidirectional Transmittance Distribution function (**BTDF**). The BRDF models reflections and diffuse scattering towards the backward half space. The BTDF models refraction and scattering towards the forward half space. While the BTDF can technically also account for absorption, the light attenuation of this phenomenon is typically covered by an absorption coefficient.

The BRDF is the central model of opaque material reflectance. It shall therefore be discussed further in section 2.4, including mathematical background.

2.3.4. Physically Based Rendering

The output medium of computer graphics is usually a screen, which cannot display the full dynamic range of the real world, and later consumed by a human observer. Therefore, traditional rendering concepts focus on perceptional realism towards the human observer on a low dynamic range screen and not physical realism of the virtual camera output. By coupling the rendering to human perception, non-physical light units, low dynamic range imaging, as well as empirical lighting and material models, are often used. As a result, material textures in traditional rendering concepts often include lighting information and disobey physical laws.

Rendering research has mostly focused on the light transport problem itself [67] as well as the material models [20]. Monte-Carlo methods can be used to get an unbiased solution of the light transport problem (cf. Section 2.3.2). Physically accurate microfacet BSDF models can be used for light matter interaction (cf. Section 2.3.3 and 2.4). However, a physically realistic image can only be simulated when the scene description itself is physically plausible. This includes material models and textures, light sources, and cameras.

Physically based rendering (PBR) is a concept that was introduced in 2004 by Matt Pharr, Greg Humphreys, and Pat Hanrahan and implemented in their scientific renderer PBRT [118]. They propose to strictly separate lighting and material models, to only use models that follow physical rules, and to correlate the model parameters with measurements and high fidelity simulations. More precisely, this includes using radiometric light units and high dynamic range imaging, spectral color representations, Monte-Carlo light transport algorithms, and material models that follow energy conservation, Helmholtz reciprocity, and the Fresnel law. By decoupling the rendering from human perception, the accuracy of the renderings is increased for arbitrary lighting, material, and scene compositions. Scene creation is also simplified, by reducing the degree of freedom of the models and textures through referenced parameters.

The concept was initially implemented 2004 with the PBRT scientific renderer [118], and the Mitsuba Render [65], which I used for my work in Chapter 4 and Chapter 5. Offline production rendering systems have also adopted PBR concepts, *e.g.*, for visual effects or design and visualization, and more recently by

real-time rendering systems alike.

A principled PBR shading model was introduced by Burley *et al.* in 2012 [15]. They propose an artist-friendly BSDF model that is based on microfacet theory 2.4. It can be used in conjunction with high dynamic range imaging, physically correct light transport algorithms, and PBR lighting to create high-quality renderings. Due to its simplicity and physically accurate result, this model has elicited a recent trend of the computer graphics industry to move towards PBR rendering approaches. Starting in 2013 [97], the progress in this area has been well documented in the popular workshop series *Physically Based Shading in Theory and Practice* at the computer graphics conference Siggraph.

Nowadays, almost all offline rendering tools, *e.g.*, Blender [11], and real-time engines, *e.g.*, the Unreal [68] and Frostbite Engines [77], feature some aspects of PBR.

Due to the emergence of physically based rendering systems, state-of-the-art texturing tools and texture assets have lately been designed for physically based material workflows, *e.g.*, using Substance Designer[1] or Quixel Studio [2].

2.4. Opaque Material Reflectance

Opaque materials are mostly impenetrable to electromagnetic waves. Conductive materials, such as metals, are always impenetrable to electromagnetic waves and, therefore, always opaque. For dielectric materials, the penetration depth of electromagnetic waves depends on material properties and the wavelength, *i.e.*, a material that appears to be opaque to light waves can be translucent to larger wavelengths.

There is no clear definition or threshold of the attenuation coefficient for when a material can be considered mostly impenetrable. However, for the context of optical simulations in computer graphics, I define opaque materials as all materials, whose average light wave penetration depth is smaller than the spatial resolution of the material models. Materials which get penetrated by light waves further than the spatial resolution, but are not yet translucent, I define as opaque materials with a significant amount of subsurface scattering.

The BSDF models the reflection, scattering, and transmission properties of arbitrary materials, *c.f.*, Section 2.3.1. It consists

[1] www.substance3d.com
[2] www.quixel.com

of a Bidirectional Reflectance Distribution Function(**BRDF**) a Bidirectional Transmittance Distribution Function(**BTDF**). Opaque materials can be modeled with a BRDF only. Opaque materials with subsurface scattering can be modeled with a modified BRDF (*c.f.*, Figures 2.4 and 2.5), the Bidirectional Subsurface Scattering Reflectance Distribution Function (**BSSRDF**) (*c.f.*, Figure 2.6). In the following section, I detail the functions and their physical properties.

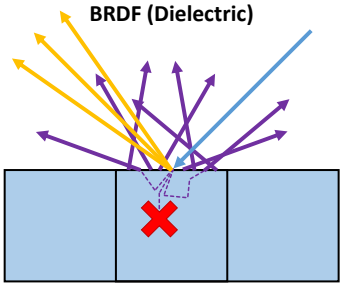

Figure 2.4.: Light, that is not reflected at the top layer of the surface, can penetrate opaque dielectric materials. Once inside the material, the light is scattered back outside the surface or absorbed. The BRDF models the reflection and scattering of light. It is a material specific reflectance function, that depends on the incident direction of the light, the viewing direction, and the wavelength of light.

2.4.1. BRDF

The BRDF is a material specific function and at the very core of modeling optical properties of opaque surfaces, along with the surface normal. It was introduced by Fred Nicodemus in [112] and describes how a surface reflects and scatters incident light, by means of a reflectance distribution:

$$f_r(\omega_i, \omega_o) = \frac{dL_o(\omega_o)}{dE_i(\omega_i)} = \frac{dL_o(\omega_o)}{L_i(\omega_i)cos(\Theta_i)d\omega_i} \tag{2.5}$$

where $L_o(\omega_o)$ is the radiance reflected towards the outgoing direction ω_o as in equation 2.4 and $E_i(\omega_i)$ the incident irradiance along the incoming direction ω_i.

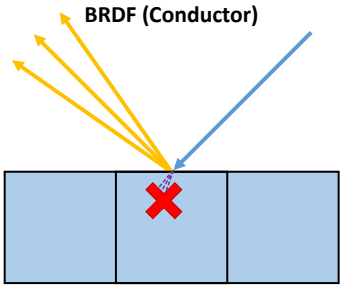

Figure 2.5.: Light cannot penetrate metallic surfaces. A large portion is reflected at the top layer. The rest is directly absorbed. As for opaque dielectric materials, the reflectance properties of metals are modeled with the BRDF.

Some principles have to be ensured to make BRDFs physically plausible. Followingly, the mathematical requirements of the principles shall be listed, and their implications for light transport simulations shall be assessed. The BRDF is required to be positive:

$$f_r(\omega_i, \omega_o) \geq 0 \tag{2.6}$$

This implies that an opaque surface cannot reflect or scatter negative radiance.

In addition, a BRDF has to follow the principle of Helmholtz reciprocity:

$$f_r(\omega_i, \omega_o) = f_r(\omega_o, \omega_i) \tag{2.7}$$

It means that the light path can be reversed without changing the flux of light. It is a central requirement for physical light transport simulation, as most light transport algorithms trace light in reverse for optimal performance. However, this principle is often violated by rendering systems, that bake precomputed lighting information into material textures.

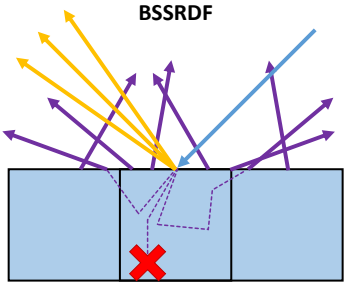

Figure 2.6.: Some mostly opaque dielectric materials possess notable subsurface scattering. Light can penetrate the material deeply, getting scattered far away from the entry point. This phenomenon can be modeled with the BSSRDF, where the reflectance of the surface depends on the incident and viewing directions as well as the corresponding surface positions.

Furthermore, a BRDF is required to be energy conserving:

$$\int_{\Omega} f_r(\omega_i, \omega_o)(\omega_i n) d\omega_o \leq 1 \tag{2.8}$$

which means that no additional radiance is created when light is reflected or scattered. It should be noted, that the BRDF accounts for light absorption. Hence, $\int_{\Omega} f_r < 1$ is physically plausible. Energy conservation is a principle that is violated by some BRDF models, such as the empirical Phong model [119].

SVBRDF In practice the 7-dimensional spatially varying spectral BRDF (SVBRDF) forms the basis for most material models in computer graphics:

$$f_r(\omega_i, \omega_o, x, \gamma) = \frac{dL_o(\omega_o, x, \gamma)}{dE_i(\omega_i, x, \gamma)} \tag{2.9}$$

where the reflectance distribution of the material depends on the surface position x and the wavelength γ. For clarity, I refrain from using the terms SVBRDF or spectral SVBRDF whenever possible, as most computer graphics research use the term BRDF as a synonym for the spectral SVBRDF. Unless stated otherwise, the term BRDF will henceforth always refer to the 7-dimensional spectral SVBRDF.

BSSRDF Some dielectric materials are subject to subsurface scattering, where incident light is penetrating the surface more deeply and scattered away from the entry point of the light. Organic materials, *e.g.*, skin and vegetation, often exhibit notable subsurface scattering. The 9-dimensional BSSRDF was introduced in [138] and models this behavior:

$$f_r(\omega_i, x_i, \omega_o, x_o, \gamma) = \frac{dL_o(\omega_o, x_o, \gamma)}{dE_i(\omega_i, x_i, \gamma)} \tag{2.10}$$

where radiance that is reflected towards ω_o at the point x_o is related to the incident irradiance from ω_i at the point x_i.

Explicit functions for BRDFs and its higher dimensional variants do not exist. BRDFs can either be sampled with measurement setups, *e.g.*, gonio-reflectometers [34] and BTF-Domes [131], or they can be approximated with analytical functions. Both methods will be highlighted in more detail in sections 2.4.2 and 2.4.4.

2.4.2. Analytical BRDF Models

Typically, analytical BRDF models are used by rendering systems to model the optical material properties. In contrast to sampled measurements, analytical BRDF models require significantly less data storage, can be used for real-time GPU rendering and produce significantly less noise with Monte-Carlo light transport algorithms due to importance sampling.

Analytical BRDF models can be split into two categories: One, classical empirical models. Second, microfacet models that are physically accurate on a geometric optics level.

Before I describe the models, I list the mathematical quantities, that are used throughout this section:

- k_d: The diffuse reflectivity
- k_s: The specular reflectivity
- ω_i: The normalized light direction
- ω_o: The normalized view direction
- $h = \frac{\omega_i + \omega_o}{||\omega_i + \omega_o||}$: The half vector or normalized bisector of view and light direction
- m: The normal of the microfacet
- n: The macroscopic surface normal

Empirical BRDF model

The **Blinn-Phong model** was introduced by Jim Blinn in [13] as a computationally efficient improvement of the original Phong model [119]. It models specular reflection as a cosine lobe and diffuse scattering as a constant lambertian term:

$$f^p(\omega_i, \omega_o) = k_d \frac{1}{\pi} + k_s \frac{n+2}{2\pi} \cos^p \Theta_h \tag{2.11}$$

where $p \in [0, \infty]$ is the specular phong exponent and Θ_h the angle between the half vector h and the surface normal n. The model is not physically plausible as it disobeys energy conservation 2.8 at grazing angles, *i.e.*, light reflected below the surface is lost. The specular term can only model isotropic reflectance and does not increase reflectivity towards grazing angles according to the Fresnel equations. Additionally, the specular exponent is a somewhat arbitrary quantity and lacks physical meaning. The

Blinn-Phong model has been a popular choice in the past, due to its low computational complexity and reasonable quality for real-time applications.

The anisotropic **Ward model** was introduced in [153]. It models specular reflection as an anisotropic Gaussian lobe and diffuse scattering as a constant Lambertian term. The original model was later improved with energy conservation at grazing angles by Geisler et al. [39]. This improved model was used throughout the thesis and reads:

$$f^w(\omega_i, \omega_o) = k_d \frac{1}{\pi} + k_s \frac{1}{\pi \alpha_x \alpha_y} \frac{\exp(-\tan^2 \delta(\frac{\cos^2 \Phi}{\alpha_x^2} + \frac{\sin^2 \Phi}{\alpha_y^2})}{4(\omega_i \cdot h)^2 (n \cdot h)^4} \tag{2.12}$$

where δ, Φ are the elevation and azimuth angle of the half vector and $\alpha_x, \alpha_y \in [0, 1]$ the roughness parameters in tangent and bitangent direction of the half vector. The main benefits opposed to the Blinn-Phong model are energy conservation and anisotropic specular reflectivity, *i.e.*, the reflectivity changes with the azimuth angle of the half vector. The roughness parameters are inspired by physical based roughness according to the microfacet theory (*c.f.*, the next Section 2.4.2). However, they lack a direct physical correlation due to the empirical nature of the model. The model does not handle fresnel reflectivity that increases towards grazing angles.

The Ward model is a computationally inexpensive anisotropic model. However, Fresnel reflectivity is deemed more relevant than anisotropy in most computer graphics applications. Hence, the Ward model has been mostly discarded in favor of physically based microfacet models. When required, anisotropic microfacet models can be used to cover all specular effects, but are computationally expensive. Subsequently, I assess the microfacet theory and its corresponding models in more detail.

Microfacet Theory and Models

Microfacet models are a class of analytical BRDF and BSDF models that are based on microfacet theory. The theory is accurate on a geometric optics scale. It models the reflection, scattering, and transmission behavior of materials with a distribution of microfacets, that are larger than the wavelength of light but not yet visible to the human eye. Microfacet theory is the core theory for material models of physically based rendering [118],

see section2.3.4.

The original microfacet BRDF model is the **Cook-Torrance** model. The basic model, alongside principles of microfacet theory, was introduced in [20]. Different modifications of the model have since been proposed. They are mostly focusing on the distribution functions of the microfacet. A generalization of the theory to BSDF models for arbitrary materials and an assessment of different distribution functions was published by Walter et al. [152]. In the following, I describe the basic Cook-Torrance BRDF model and its components. Then I outline different choices for each component of the model.

According to microfacet theory, a macro surface BRDF can be modeled as a collection of microfacets. Each microfacet reflects and refracts light into the perfect mirror and refraction directions, following the Fresnel equations. The Cook-Torrance BRDF models the reflection and scattering properties of the macro surface by means of statistical distributions. It is a compound BRDF, consisting of three terms, that each model different physical behavior:

$$f^m(\omega_i, \omega_o, \alpha) = k_d \frac{1}{\pi} + k_s \frac{F(\omega_i, h) D(h, \alpha) G(\omega_i, \omega_o, h, \alpha)}{4|\omega_i \cdot n||\omega_o \cdot n|} \quad (2.13)$$

where

- $\alpha \in [0, 1]$: Is the roughness of the material. It models the variations of the microsurface. $\alpha = 1$ corresponds to a random orientation of microfacets and thus to a very rough microsurface. $\alpha = 0$ corresponds to perfectly smooth microscopic surface.
- $F(\omega_i, h)$: Is the fresnel factor for incident light from ω_i and the half vector h.
- $D(h, \alpha)$: Is the microfacet distribution function that models the probability that microfacets are aligned along the half vector h
- $G(\omega_i, \omega_o, h, \alpha)$: Is the shadowing-masking function that models the probability that a microfacet along the half vector h is visible in both directions of the light.

The choice of the distribution functions may vary, depending on the desired properties of the resulting BRDF model. How-

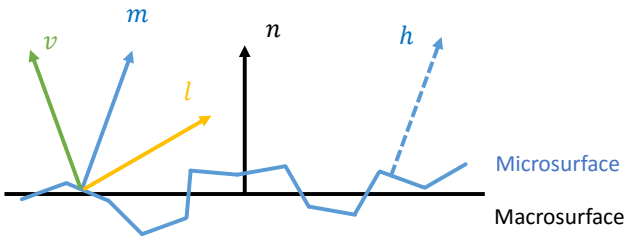

Figure 2.7.: The detailed microsurface m is replaced by a macrosurface n and a microfacet distribution function, which models the probability for light to be reflected towards the halfvector h.

ever, they have to follow certain principles to be physically accurate. I detail the properties of the distributions and the underlying principles in the following paragraphs.

Microfacet Distribution Function The statistical distribution of micro surface normals m is described by the microfacet normal distribution function $D(m)$, a density function with the unit $\frac{1}{sr}$. A plausible distribution of microsurfaces has to obey certain principles [152]:

- The density function is positive: $0 \leq D(m) \leq \infty$
- The micro surface area is not smaller than the macro surface area: $1 \leq \int D(m) d\omega_m D$
- The visible area of the micro surface is equal to the visible area of the macro surface: $\omega_o \cdot n = \int D(m)(v \cdot m) d\omega_m$

$D(h, \alpha)$ models the probability that the normals of the micro surface, that belong to of macro-surface with a roughness α, point towards the half vector h, see Figure 2.7. The distribution function for the macro surface BRDF $D(h, \alpha)$ can be derived by integrating the projected micro surfaces towards the view direction ω_o. This derivation can only be formulated in conjunction with the corresponding shadowing-masking term $G(\omega_i, \omega_o, h, \alpha)$. For extensive mathematical coverage, I refer to Walter et al. [152]

Shadowing-Masking Function The Shadowing-Masking function $G(\omega_i, \omega_o, m)$ models the fraction of the micro surface with the micro normal m, that is visible by both the light source ω_i and the viewer ω_o (See Figure 2.8). Some properties are required to make it physically plausible [152]:

- The visible area cannot be negative or larger than the total area: $0 \leq G(\omega_i, \omega_o, m) \leq 1$

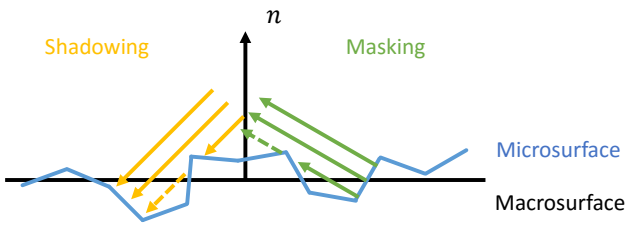

Figure 2.8.: The Shadowing-Masking function describes the ratio of the surface area that is visible in both the incident direction (yellow) and outgoing direction (green). In this example, a portion of the light is blocked in either direction and the shadow-masking terms are smaller than one.

- It is Helmholtz reciprog: $G(\omega_i, \omega_o, m) = G(\omega_o, \omega_i, m)$
- The backside of the microsurface is never visible from any direction:
  $G(\omega_i, \omega_o, m) = 0$ for $(\omega_i \cdot n)(\omega_i \cdot m) = 0$, or $(\omega_o \cdot n)(\omega_o \cdot m) = 0$

Shadowing-masking is mostly relevant for very rough surfaces and near grazing angles and needed for energy conservation. The exact formulation is rarely possible, but shadowing-masking of a macro surface can be approximated in conjunction with the selected microfacet distribution $D(h, \alpha)$. For a mathematical derivation and a detailed assessment of existing shadowing-masking functions, I refer to [152].

Choosing Microfacet Model Terms The Fresnel Equations formulate the exact solution of the **Fresnel term** based on a complex index of refraction. The Schlick Approximation [129] was developed to approximate the Fresnel term for dielectric materials efficiently. It omits the polarization of light, and therefore a real-valued index of refraction is used. Metals can be approximated with substantial errors, by substituting the index of refraction with the specular color at normal incidence. The Schlick Approximation is almost exclusively used in modern computer graphics. The exact, computationally expensive fresnel term is mostly applied in scientific rendering systems [118].

The chosen **microfacet distribution function** of the original paper by Cook and Torrance [20] is the Beckmann distribution [9]. It assumes a Gaussian distribution of the microsurface normals and is a popular choice in optics. A popular choice in computer graphics is the Phong-microfacet distribution [13], due to its computational efficiency. The Phong distribution can be modified to cover anisotropic microfacet distribution [5]. More recently, the GGX distribution has been established [152]. It was shown to best match measured data, due to its long tail.

The **shadowing-masking function** has to be derived for the selected microfacet distribution as a closed form solution. The v-cavity surface model was suggested in the original paper to guarantee energy conservation for an arbitrary microfacet distribution [20]. However, it contains discontinuities in the first derivative and should therefore not be used [152]. The Smith shadowing-masking approximation [137] was derived for Gaussian surfaces and is used almost exclusively in today's computer graphics. However, Smith integrals do not have a closed-form solution for the Phong distribution. Therefore, a new shadowing-masking function for the Phong distribution was introduced along with the Ashikhmin and Shirley model [5].

2.4.3. Bidirectional Texture Function

When approximating an object with a coarse geometry, the outgoing radiance for any illumination direction can be modeled as an 8D reflectance field [23]. Subsequently, I refer to the 9D spectral reflectance field for the sake of consistency. For light sources and observers outside of the object, novel view and light directions can then be reconstructed from existing samples, *e.g.*, from measurements, as a linear combination. When using the actual object geometry, it is equivalent to the 9D BSSRDF, yet significantly easier to sample and reconstruct.

Under the assumption of a distant light source, the reflectance field can be reduced in dimensionality to the 7D Bidirectional Texture Function(**BTF**) [22]. This assumption is, in most cases, reasonable as mesoscopic geometric details on the pixel scale are typically magnitudes smaller than the distance to the light source. In contrast to the SVBRDF of the same dimensionality, the BTF accounts for non-local light transport, *i.e.*, subsurface scattering and specular interreflections, and can be considered to be equivalent to the BSSRDF in terms of captured reflectance. However, as it models these non-local scattering events, the function no longer obeys physical principles, *e.g.*, Helmholtz reciprocity 2.7, and is therefore nearly impossible to synthesize in plausible ways.

Typically, the BTF is used to sample the reflectance field of objects in discrete steps, which are acquired from measurements, *c.f.*, Section 2.4.4. As proposed by Dana *et al.* [22], it can be stored in a set of textures that vary by discretized light and camera directions ω_i, ω_o:

$$BTF(x, y, \omega_i, \omega_o) = \{T(x, y)_{\omega_i, \omega_o}\} \tag{2.14}$$

where T is a texture with the texture coordinates $x, y \in I' \subset \mathbb{N}^2$ which are resampled from the set of the originally captured images I, and $\omega_i, \omega_o \in M'$ with M' are the light and camera directions that are resampled from the set of original camera and light directions M.

Resampling of the original captured image data I is typically performed to consolidate each captured image into a single re-projected and discretized texture space [106]. Resampling of the original camera and light directions in M is performed because of two reasons: First, the BTF texture dimension can differ from the physical set of cameras and light sources [131]. Second, the corresponding camera and light directions are different for each point on the material sample [106].

An alternative notation of the BTF is that of a set of spatially varying apparent BRDF (ABRDF), which was introduced by Wong *et al.* [160]:

$$BTF(x, y, \omega_i, \omega_o) = \{A(\omega_i, \omega_o)_{x,y}\} \tag{2.15}$$

where A is the ABRDF that varies for the texture points x, y and depends on the light and camera directions ω_i, ω_o. Typically, it contains the factor $\omega_i \cdot n$ with n the coarsely scanned surface normal, as the measured reflectance data depends on the incident irradiance, which is attenuated by the angle of the light. The ABRDF for a surface point x, y can be normalized with the incident irradiance on its captured surface normal n:

$$\hat{A}(\omega_i, \omega_o)_{x,y} = \frac{A(\omega_i, \omega_o)_{x,y}}{\omega_i \cdot n} \tag{2.16}$$

in this format, the ABRDF is equivalent to a discretized sampling of the continuous BRDF with the additional inclusion of non-local scattering effects.

2.4.4. Capturing opaque material reflectance

While there are a lot of analytical reflectance models available, only a few publications are concerned with measurements of reflectance parameters for opaque surfaces. Measuring and storing highly dimensional reflectance data is a complex task. Usually, some compromises in terms of spatial, spectral, or angular accuracy have to be made to facilitate a feasible acquisition pro-

cess and data storage. Standard approaches that can measure BRDF and in some cases SVBRDF include setups with moveable sensors and light sources and in some cases mirrors to increase the number of measurement samples taken at a time. Nowadays, some setups even allow capturing approximations of the reflectance fields, which are an equivalent to a full 9D BSSRDF.

BRDF and SVBRDF Acquisition

The first measurement setups for SVBRDF and BRDF data were gonio-reflectometers, such as the one described in [34]. They consist of a movable light source, a detector, and usually a planar material sample. The detector is in most cases a spectro-radiometer, allowing a very dense sampling of the light spectrum. The drawback of such setups is, that even though the spectral resolution is high, it is not a feasible approach for the capturing reflectance data with high spatial or angular resolutions since only one sample is acquired at a time.

This process was sped up significantly by the introduction of the imaging gonio-reflectometer setup of Ward[153]. Here the spectro-radiometer is replaced by a CCD camera with a fish-eye lens and a hemispherical mirror. This allows multiple measurements at a time and thus an increased angular and spatial accuracy as well as anisotropic reflectance within a reasonable timeframe. The drawback of having an RGB-binned spectrum does not weight too high, as most computer graphics applications apply RGB-binning aswell.

Another approach was introduced by by [89] and later refined by [93]. Here, instead of using a mirror, a spherical material sample is used that allows the acquisition of dense BRDF samples. A fixed camera is used with a moveable light source. Measurement noise is filtered by having multiple measurements for identical light-normal-view setups. Thus, the system is very fast and not very error prone. The approach by Matusik [93] later increased the sampling density of the BRDF($4M$ samples per material). However, it is limited to acquiring isotropic BRDF data without spatial variations. The acquired reflectance data was published in the MERL BRDF database, which is commonly used in computer graphics.

Reflectance Field Acquisition

When neglecting phosphorescence and fluorescence, the scattering of light on a surface can be described by the 9-dimensional BSSRDF[138]. If instead of the actual surface an approximation, *e.g.*, a coarser polygon model, is used, the reflectance field [23] describes precisely the same appearance and can be used as a substitute.

Acquiring full reflectance fields is time-consuming and expensive. Storing the data takes up a lot of disk space. Consequently, to the best of our knowledge, only fractions of the function's 9D parameter space have been sampled extensively or simplifying models have been employed to reduce the dimensionality. The spectral accuracy, the spatial accuracy, and the angular accuracy all have to be taken into consideration.

One especially powerful simplification that, in practice, yields high-quality results is the Bidirectional Texture Function (**BTF**). In contrast to the full reflectance field, the BTF makes the simplifying assumption that the incoming illumination comes from distant light sources. This eliminates the dependency on the position the light enters the surface. Hence, the dimensionality is reduced to 7D, allowing an efficient and sufficiently dense sampling. Although the far-field illumination assumption is in reality not exactly fulfilled, the distances to light sources are usually several orders of magnitude larger than the tiny structures that make up the appearance of the surface. At least for small, localized patches on the surface, distant illumination is, therefore, a valid approximation. Hence, the BTF performs very well for the reproduction of materials that are opaque or do exhibit localized sub-surface scattering [105]. For these cases, the relaxation to far-field illumination is a much lower restriction than the model assumptions that are found in the also prevalent 7D SVBRDF representation.

Nowadays, mostly camera array-based approaches are used to measure BTFs. One powerful measurement setup with 151 light sources and cameras and a structured light scanner is the DOME II setup of the University of Bonn [131]. It combines the acquisition of BTF data, which is stored in an assembly of RGB textures, with the acquisition of the surface geometry, which is stored in a mesh. Other setups for measuring BTFs without geometry are gonio-reflectometers such as the setup of the original BTF paper [22] and mirror based setups [53].

2.5. Wet Surfaces

In this section, I outline the theoretical backgrounds of the reflectance properties of wet surfaces. I identify two physical phenomena that cause a change in the appearance of wet surfaces: First, water saturation of pores within the material. Second, a water layer on top of the material surface. In the next part of this section, I detail a physically based wet surface model that can be applied to an image-based augmentation setup, *e.g.*, the work in Chapter 6.

2.5.1. Observation and theory

When comparing wet and dry surfaces, the primary visual differences that one can retain is that wet surfaces look darker and more reflective.

The extent of both phenomena differs by material type. It is very apparent for: i) Powdered materials, *e.g.*, sand and soil. ii) Porous compound materials, *e.g.*, concrete, asphalt, stone, and wood. iii) Absorbent materials, *e.g.*, fabrics and paper. By contrast, there are only very subtle appearance changes for smooth surfaces, *e.g.*, glass, metal, and smooth coated materials. In general, rough surfaces seem to be more heavily affected than smooth surfaces. According to extensive studies about the wetting process and its effect on optical material properties [145, 111], this is explained by two physical causes: First, most rough materials have small pores and air gaps within the material. Once the pores fill up with water, the optical material properties change. Second, a thin layer of water accumulates on the material surface, causing additional changes in the optical properties.

When the water cannot be drained or seep through the ground surface, *e.g.*, because it is already saturated, the initially thin layer of water, which sticks to the ground surface, starts to accumulate and flow to height minima of the ground surface.

Dielectric porous materials are profoundly affected by the water saturation of pores and the layer of water alike. Smooth dielectric, with less and smaller pores, and metal are mostly only affected by the layer of water. In the following, I detail optical effects from a water saturation of pores and a layer of water concerning the formation and the influence on optical material properties.

Water Saturation of Porous Dielectrics

The findings of this section mostly refer to the work presented in [145, 111]. In other cases, I refer to the corresponding papers.

However, first, I clarify what porous materials are to understand its interaction with water better. Surface roughness refers to micro-geometric features on a scale that is larger than the wavelength of light, yet not directly visible by the human eye. 2.4.2. Those micro-geometric features directly influence light-matter interaction, thus changing how light is refracted and reflected. Almost all natural and human-made materials are rough to some extent. Porosity is often involved in natural materials, *e.g.*, sand, stone, clay, and almost always involved in industrially produced materials, that are based on powders, *e.g.*, asphalt, concrete. It is different from roughness in the way that it refers to holes, pores, and air cracks that exist within the material.

When interacting with dielectric materials, light is either reflected or refracted into the material surface. Within the material, a portion of the light is absorbed, and the remaining portion is either transmitted or scattered. The light that is scattered back to the surface is what the observer perceives as Lambertian reflectance. Within porous materials, air-material interfaces exist at each of the pores. At these pore-interfaces, light is refracted and reflected according to Snell's law.

The pore-interfaces change when the air, with a refractive index (IOR) of 1.0, is replaced by water, with an IOR of 1.33. Due to a decrease in the difference of IORs, less light is reflected at the pores, and the light is refracted to a lesser extent. The travel of light is more forward and, thus, it needs to travel longer before it can make its way back to the surface. Due to the longer light paths within the materials, the likelihood that the light is absorbed increases and the diffusely scattered appearance becomes darker.

The water saturation of the pores causes an additional reflectance effect. The reflectivity of the material is increased even when no water is covering the surface. This can be explained. When the pores fill with water, a large portion of the reflected light comes from the water instead of the material. As water is almost perfectly smooth, the specular reflectance of the material becomes more focused and sharp. This behavior is similar to that of a material coating, *e.g.*, from a polished wooden floor.

However, another factor comes into play. At each pore-interface, a portion of the light is reflected backward. The most significant portion of such reflected light comes deep layers, is subsequently scattered, and therefore forms a broad specular distribution. When the air is replaced by water, less light is reflected by the pore interfaces due to the change of IORs and, consequently, less broadly distributed specular light paths contribute to overall specular reflectivity. Essentially, this means when a porous material becomes water saturated, overall less light is specularly reflected but more of it along the perfect mirror direction.

According to [79, 85], the darkening of wet surfaces is accompanied by a notable change in hue and saturation.

Layer of Water

At air-water interfaces, some light is reflected, and the rest is transmitted and refracted. The reflected portion is 2% at normal incidence and increases to 100% towards low grazing angles. The directions of the reflection and refraction depend on the thickness of the layer. For thin layers of water, they follow the underlying surface geometry, whereas, for thick layers of water, the directions are smoothly aligned. As a result, puddle reflections appear sharp and thin layer reflections appear more scattered and diffused-

The refracted portion of the light is transmitted towards the underlying surface. While traveling inside the water, the light may be scattered or absorbed. At the underlying surface, it can be reflected and, subsequently, travels back to the water-air interface (this time in reverse). However, depending on the exit angle on the interface, total internal reflection may occur. Subsequently, the light can travel back and forth between the surface and the interface, being subjected to additional rounds of absorption.

The absorption of light depends on the absorption coefficient α and the traveled distance. Integrated over the entire visible spectrum the absorption coefficent is $\alpha \approx 0.12m^{-1}$ [121]. However, in reality, it is most abundant near the red spectrum and almost zero at the blue (hence the blue taint of large water bodies). For a thin layer of water, the absorption can always be neglected, as the traveled distances are magnitudes smaller than one meter. For a thick layer of water, the absorption increases but remains barely notable for reasonable accumulation

levels of water puddles. By contrast, the absorption increases for total internal reflections. Assuming a perfectly smooth underlying surface, a darkening can be observed below the critical angle. However, for a rough underlying surface, light is diffused towards all directions, and some darkening from a layer of water can be observed in all directions.

The scattering of light traveling within the water causes light paths to exit the surface in arbitrary directions and distant locations. As a result, the appearance of the underlying surface is diffused, an effect that increases with the thickness of the water layer.

2.5.2. Physically Based Wet Surfaces

The effects from a layer of water can be modeled with layer BRDF models, *e.g.*, the Weidlich and Wilkie layered BRDF model [154]. Here, the Fresnel terms are evaluated at both layers, and the light paths attenuated from water absorption. However, this double layer BRDF models is computationally expensive, cannot be evaluated for image-based lighting, and is therefore not applicable to the problem in Chapter 6. Additionally, the model does not model the water saturation of pores and is, thus, not suited to simulate the entirety of ground surface wetness effects.

Lagarde presents an alternative approach in an online survey [76]. The author analyzed existing models for wet porous material by Merillou [104] and Hnat *et al. et al.* [59]. He referenced the models with measured data of wet ground surfaces and encoded all reflectance effects in a single layer physically based wet surface model based on the BRDF format of Disney [15]. This allows the dynamic interpolation of dry and water saturated states of porous dielectric materials. The accumulated water layer is modeled by a surface normal and BRDF interpolation with a smooth water material. This approach is limited to scattering events inside the wet porous material and from its top layer. Total internal reflection and absorption from the water layer cannot be modeled (*c.f.*, Section 2.5.1

However, according to the study by Weidlich and Wilkie[155], both effects are negligible when the layer of the water is thin, and the difference in IOR is minimal. Typically, total internal reflections amount to an overall attenuation factor of $0.95 - 1.0$ and absorption to an attenuation factor of $0.98 - 1.0$ for

accumulated water layers at urban ground surfaces.

Subsequently, I present Lagarde's [76] derivation of the physically based wetness model based on dry and wet ground surface measurements.

Two studies based on measurements aimed to find a function for the darkening of wet surfaces are [79] and [145]. They show that the wet albedo is a non-linear function of the index of refraction of the surface and water as well as the dry albedo color. Medium albedos (between 0.2 and 0.5) are reduced more by wetting than higher and lower albedos. They explain that for lower albedos, more light is absorbed on initial contact with the surface. For higher albedos, more light is directly reflected. In both cases, less light is reflected internally, where the darkening occurs. It should be noted, that there are some differences between the two models. The wet albedo function of Lekner et al. [79] exhibits less darkening for lower albedo values than the wet albedo function of Twomey *et al.* [145]. The functions are fitted to different measurements. Consequently, [79] seems to work better for rough solid surfaces such as asphalt and concrete and [145] is best applied to finely divided materials, such as sand.

Lagarde [76] criticizes that the function is only dependent on IOR and albedo values, completely neglecting the impact of roughness and porosity properties. However, roughness and porosity are a major factor, *c.f.*, Section 2.5.1.

The study from [167] highlights that the Lekner and Twomey models should be taken into account in addition to microscopic roughness and a ratio for translucent particles, which is a concept similar to that of porosity. They state that the concentration of translucent particles causes the difference between dry and wet reflectance. Additionally, rough surfaces scatter light more diffusely even when a thin layer of water is present.

Two studies conducted BRDF measurements of dry and wet rough dielectric materials [139], [45]. A Cook-Torrance microfacet BRDF was fitted to the measurements, which gives us a diffuse color, a specular color and a roughness term for each surface type and both states. The resulting models are thus dependent on dry albedo, roughness, and material type. Unfortunately, porosity is not considered. A spatially constant porosity factor is implicitly included from the measurements, but no generalization for other materials or a spatially varying porosity can

be derived.

The proposed model for physically based wet surfaces by Lagarde [76] utilizes the measurement and models [139], [45] and generalizes the model parameters for rough, porous dielectric materials. The author proposes to extend the model with an attenuation factor ρ based on the effective porosity similar to Hnat *et al.* [59] and Merillou *et al.* [104]:

$$\rho = lerp(0.2, 1.0, p) \tag{2.17}$$

where $lerp()$ is the linear interpolation function with the third parameter as the blending weight, and p is the porosity attenuation factor. The lower bound $\rho = 0.2$ based on the ranges found in [45] and the empirical values in [108]. The effective porosity can be controlled by artists as a porosity texture or derived as a constant value from literature [59] [104]. In either case, the author suggests porosity values of $\alpha \approx 0.5 - 0.7$ for concrete and asphalt to match measurements from [139], [45]. Using this attenuation factor, the final model of the physically based wet surfaces can be formulated:

$$f_{wet} = lerp(k_d, \rho * k_d, w) f_d + k_s f_s(lerp(\alpha, \alpha * \rho, 0.5 * w)) \tag{2.18}$$

where $lerp()$ is the linear interpolation function with the third parameter as the blending weight, w is the water saturation of wetness level, f_d, f_s are the diffuse and specular terms of the physically based Disney model [15], and α is the roughness parameter of the specular BRDF term.

3

Related Work

SYNTHETIC datasets frequently come to operation for developing and testing computer vision algorithm and modules. They are mostly used for computer vision performance analysis due to the availability of accurate ground truth. However, systematic differences exist between the performance of computer vision algorithm on real and synthetic data. As a result, performance assertions based on synthetic datasets alone often do not translate well to real-world datasets. With the recent emergence of Deep Neural Networks (DNNs) for many computer vision tasks, large amounts of data are needed for training and testing. Computer graphics can be used to generate synthetic training data quickly. However, DNNs solely training on synthetic data, often perform poorly on real datasets. Better network performance can be achieved by mixing synthetic datasets with real-world datasets during training. An alternative training approach that often leads to excellent network performance is the augmentation of real-world datasets.

In the following sections, I outline the published work of the topics above regarding *low-level vision*, *high-level vision*, *systematic differences to real-world data*, and *augmentation of real-world data*. More specifically, Section 3.1 is concerned with synthetic datasets for low-level vision tasks, such as optical flow and disparity estimation. Synthetic datasets are essential for this class of algorithms, as accurate real-world reference and training data is costly, hard to acquire, and inherently inaccurate. Section 3.2 outlines recent progress in using synthetic datasets for high-level vision tasks, such as segmentation and object detection. Synthetic datasets are a promising tool for these tasks, as state-of-the-art algorithms are neural network based with a substantial need for training data. Only a few studies have been published, that specifically addresses the systematic differences between real-world and synthetic datasets. They are presented in Section 3.3. Section 3.4 outlines the work on two augmen-

tation approaches. This includes data augmentation methods, *i.e.*, methods that manipulate the original data by applying phenomenological models, and scene augmentation methods, *i.e.*, methods that perform a context-aware manipulation of the scene in 3D world space.

3.1. Synthetic Datasets for Low Level Vision

To my best knowledge, the first published synthetic dataset for computer vision tasks is the Yosemite sequence [55]. It was initially created to evaluate a newly proposed optical flow algorithm. The rendering system performs basic rasterization with simplistic Lambertian materials. Therefore, the realism of the sequences is minimal.

The well-known Middlebury dataset [6] for optical flow and stereo estimation consist of two sequence types: First, laboratory sequences of rigid objects captured with stop motion techniques. Second, multiple synthetic sequences of a virtual urban and grove scene. Similar to the Yosemite sequence, the synthetic images are rendered with low realism via rasterization and simple Lambertian materials.

McCane *et al.* [98] published a benchmark suite for optical flow algorithms. The suite includes real-world and synthetic sequences. The real-world sequences are created from laboratory scenes, where the ground truth is measured by projecting the planar surfaces of the objects. The synthetic images consist of rigid objects in front of a textured background, where the virtual camera and objects are moved with varying patterns. The images, as well as the flow ground truth, are rendered with ray-casting, a ray-tracing technique where only the primary camera ray is cast into the scene. The pixel is then shaded with Lambertian reflectance. In contrast to rasterization techniques, where the ground truth is computed per vertex, the ray-casting ground truth is pixel accurate. However, the realism of the images is identical to the previously mentioned datasets.

Entire frameworks for creating synthetic test data with ground truth also exist. They include the framework of the UCL Ground Truth Optical Flow Dataset by Aodha *et al.* [4], as well as the frameworks by Haltakov *et al.* [51] and Onkarappa and Sappa [113], which both create automotive sequences with depth, flow, and segmentation ground truth. Using these frameworks, the

user can rapidly create large amounts of image and ground truth data. The synthetic data generation approach of the UCL framework is described in their previous paper [87]. They use ray-tracing combined with a Lambertian BRDF, allowing an accurate simulation of shadows and global illumination but only diffuse reflectance effects. The framework by Onkrappa *et al.* follows the same synthesis approach. The Haltakov framework targets real-time testing and, therefore, simple rasterization rendering and Lambertian BRDF models are utilized. Consequently, neither shadows nor reflectance effects are simulated accurately. All three frameworks use simplistic 3D assets and textures, and the overall realism of the simulations is low.

A work supplying a purely synthetic dataset, aimed at advanced learning-based algorithms, is the "naturalistic" Sintel dataset [16]. It consists of a large amount of training and test images along with optical flow, stereo, segmentation, and camera pose ground truth. The dataset improves upon previously mentioned datasets in terms of realism by adding thin-lens camera effects, *i.e.*, motion blur and defocus blur, and specular highlights. However, the BRDF textures remain artistically driven, lack high-frequency details, and the specular term of the BRDF model physically implausible. As a result, the specular highlights lack spatial variations as well as increased Fresnel reflectivity at grazing angles. Nevertheless, the study shows that, compared to previous datasets, the statistics of the synthetic flow fields better match real-world flow fields. These results indicate that modeling of specular reflectance effects is essential in the context of optical flow training and testing. Further improvements to the reflectance models and textures can, therefore, increase the confidence of the synthetic data.

A synthetic stereo dataset, focusing on improved realism, was published by Matrull[92]. It is a synthetic replica of the well-known *Head and Lamp* stereo dataset [109]. The images are rendered with ray-tracing and a presumably simple Phong BRDF with diffuse textures derived from the real-world photographs and artistic modeling of the specular term. The dataset includes 1800 stereo image pairs under varying lighting conditions and comes with disparity ground truth and occlusion masks. The realism of the dataset appears to be high, but no comparison to real-world datasets is conducted.

More recently, the Flying Chair dataset, consisting of more

than 20000 image pairs, was supplied for training and testing optical DNNs by Dosovitskiy *et al.* [26]. They augmented real-world backgrounds with pre-rendered synthetic images of chairs. The background image, chair models and transformations, as well as the affine motions of the camera and chairs, are all sampled in a randomized fashion. Their FlowNet optical flow estimation network was trained on the Flying Chairs dataset and performed similar compared to state-of-the-art classical optical flow algorithms and networks trained on the more realistic MPI Sintel dataset [16]. The study claims that a realistic scene composition is not relevant for training and that data variety and data augmentation are more relevant for the performance of optical networks trained on synthetic images. In a follow-up study by Ilg *et al.* [64], it became apparent that networks trained on the Flying Chairs dataset perform unreasonable under real-world conditions, *i.e.*, when small motions are present. Therefore, they propose to expand the Flying Chairs dataset with the ChairsSDHom dataset, consisting of image-pairs with smaller motions. The Flow-Net 2.0, which was trained on Flying Chairs as well as ChairsSDHom, improves the performance under realistic conditions and when small motions are present. Due to the lack of 3D-motion and lighting interaction between objects and backgrounds, both datasets are not able to simulate critical reflectance effects, *i.e.*, reflections and specular highlights. As a result, the trained networks are volatile under those real-world conditions.

A similar dataset, called Flying Things, was introduced by Mayer *et al.* [94]. It expands on the idea of creating randomized scenes for optical flow, disparity, and scene flow training data. In contrast to the Flying Chairs dataset, it features structured backgrounds and true 3D-motion of objects and cameras. The static background is composed of randomly aligned simple geometric shapes, to which a texture is applied. The camera and the more complex foreground objects follow linear paths in 3D space. As a result, networks trained on this dataset perform better for more complex motions, *i.e.*, 3D rotations and translation along the z-axis of the camera. The dataset was path traced, creating realistic lighting, but only simple material models and textures are used. Critical reflectance effects are, therefore, not simulated and tested in this study.

A visually more realistic dataset for disparity estimation un-

der difficult lighting conditions is the *UnrealStereo* dataset [168]. They utilize the state-of-the-art Unreal Game Engine and publically available assets. The assets use a physically based material model that allows accurate specular and transparency effects. The Unreal Engine approximates light transport through rasterization and sophisticated screen-space global lighting techniques. By varying the severity of the reflectance effects, they could identify that specular and transparency effects are critical to the robustness of disparity estimation methods, and therefore needed in a discriminating stereo dataset. The confidence of the dataset could be further increased by using accurate, unbiased light transport algorithms in addition to measured material textures to better simulate those critical effects. This approach was chosen for my conducted studies on reflectance model accuracy in Chapters 4, and Chapters 5.

3.2. Synthetic Datasets for High Level Vision

A virtual replica of five urban scenes of the commonly used KITTI-dataset [38] was introduced as Virtual KITTI [35]. Another synthetic dataset for the autonomous driving domain is the SYNTHIA dataset [126]. It contains a collection of urban sequences with pixel level segmentation and class annotations. Both Virtual KITTI and SYNTHIA are rendered with the Unity Engine for real-time applications. The engine uses rasterization techniques and screen-space approximations of global light paths to simulate light transport within the scenes. A physically based material model is used to model reflectance and translucency effects. The Virtual Kitti paper utilizes publicly available assets and textures, created by artists. Unfortunately, the SYNTHIA paper does not state the creation process of the material textures and assets. However, based on observation, I assume that similar artistic assets and textures are used. Both Virtual KITTI and SYNTHIA claim to be photo-realistic. However, photo-realism is not well-defined and in the eye of the beholder. With the used rendering system and material textures, physical realism cannot be achieved.

The open-source CARLA simulator, with a similar degree of realism, was released by Dosovitskiy *et al.* [27]. The simulator facilitates synthetic images and ground truth data for training and testing segmentation, disparity, and object detection networks.

Further, it can even be used for planning and driving policy tasks in a closed-loop simulation environment. The simulator is implemented with the Unreal Engine, which uses rasterization and screen-space global lighting techniques. The textures for the physically based material model are low weight textures and artistically driven. As a result, the rendered images are unrealistic and exhibit periodic texture patterns, which can be problematic for correspondence algorithms.

Shotton *et al.* made a case for purely synthetic data in the field of pose estimation [136]. They used motion capture data to create a large synthetic dataset for the training of human pose estimators. Their approach enabled them to create countless texture and color variations, *e.g.*, for human clothing and appearance.

Renderings of synthetic objects in front of real-world backgrounds are frequently used to train networks for high-level vision tasks. Pishchulin *et al.* were the first to utilize this approach [120]. To train pedestrian detectors, they used synthetically rendered humans and animations from motion capture data and placed them in front of urban background images. The domain randomization approach by Tremblay *et al.* [143] further extends this idea. They randomly place cars with arbitrary, but accurate physically based materials in front of real-world backgrounds. They show that object detectors trained on such images, perform better than those, trained on the structured, but less realistically rendered scenes of Virtual KITTI [35]. This finding is relevant for the context of this thesis, as it indicates that for neural network training, accurate material models and realistic rendering are more relevant than a semantically plausible scene composition.

Prakash *et al.* adapt the idea of domain randomization with structured scene conditions *et al.* [123]. The scenes are randomly composed according to realistic and structured automotive conditions, *e.g.*, cars are placed on road lanes and pedestrians on sidewalks. Additionally, the physically based scene materials are textured semantically correct, *e.g.*, cars are painted with car paints and roads with tarmac materials. The images are then rendered with the Unreal Engine, using rasterization and screen-space global light approximations. They can show, that high fidelity material models combined with semantically realistic textures produce synthetic datasets with the highest confi-

dence.

Tsirikoglou *et al.* published an advanced method for creating high-fidelity synthetic training and testing data [144]. They present a procedural approach to render still images of realistic urban environments with ground truth for semantic segmentation. For each image, a unique virtual world is randomly created, consisting of procedural buildings, road networks, and vegetation as well as a fixed set of weather conditions, people, and cars. In contrast to single world approaches of most other datasets, a virtually infinite amount of image variations can be created. They utilize unbiased path-tracing, physically based procedural materials, and physically accurate lighting models. The approach enables high-quality renderings without the need for manual fine-tuning of light sources and materials, which could not be applied in a procedural approach. They could show, that their high-fidelity data vastly improves semantic segmentation network performance in comparison to the low-fidelity data of SYNTHIA [126], which uses artistically created textures and biased rasterization techniques.

In the follow-up paper, this approach is used to create the Synscapes dataset with semantic segmentation and object detection ground truth in the Cityscapes and KITTI formats [161]. The fidelity of the rendered images is further enhanced by modeling camera stray light and glare effects by means of a Point-Spread-Function. They could show, that compared to training on the real-world KITTI data, almost identical performance levels can be reached with an object detector trained on Synscapes. Similar but slightly worse results can be achieved for semantic segmentation compared to the Cityscapes dataset. In both cases, a mixed training surpasses the performance of training with real-world alone.

3.3. Systematic Differences between Real-World and Synthetic Datasets

Vaudrey *et al.* [147] studied the difference between engineered sequences, *i.e.*, synthetic and laboratory sequences, and real-world sequences for low-level vision algorithms in the context of driver assistance systems. They showed that optical flow algorithms are typically more accurate on synthetic image data compared to real-world applications. They explain that in real-world data ob-

ject and texture boundaries are much less distinguished, and the brightness constancy assumption is violated more often. Both factors have a significant influence on the resulting flow fields, but most synthetic performance benchmarks do not emphasize these aspects well enough.

A similar comparison between real and realistically rendered data was carried out by Meister *et al.* [101]. The authors carefully created a real-world scene consisting of a rotating wooden block as well as a virtual replica in Blender. They used the ground truth generated by Blender to compare optical flow results between both sequences. By varying the rendering realism, they could show that illumination and material differences cause the most significant differences between flow fields from real and synthetic images.

Haeusler *et al.* [49] argue that simplistic scenes are best suited to isolate individual effects such as brightness changes and depth boundaries. They approach the problem by asking whether we can already draw essential conclusions about algorithm performance from extremely simple test images. The advantage of this approach is that vast amounts of data can be created rapidly and cost-effectively to study isolated effects to great detail.

A study regarding rendering realism and its effect on training deep neural networks for automotive object detection was carried out by Movshovitz-Attias *et al.* [103]. They show that synthetically rendered cars can generally be used to train deep neural networks for car detection. They further show that the degree of realism of the lighting conditions and the material properties directly affects the performance of the network. It is, therefore, optimal to use material models and textures that most closely match the appearance of the real-world materials.

An in-depth study regarding the quality of synthetic datasets for dense correspondence problems, *i.e.*, optical flow and stereo, was carried out by Mayer *et al.* [95]. They identify four relevant rendering aspects for the quality of synthetic datasets: *Camera models*, *textures*, *lighting* and *shape*. They state that a valid *camera model* should include two aspects: First, radial blur, which results from rectification of wide lens distortion. Second, artifacts that result from demosaicing Bayer patterns. Interestingly enough, more noticeable blur effects from the camera model, *i.e.*, optical aberrations and motion blur, are not studied. They further show that realistic *textures* with large variety

are essential for training data quality and network performance. However, they do not state which material model was fed with the textures so that a simple Lambertian model can be assumed. For the *lighting* of the scene, which in their cases also includes the reflectance models, they show that the training data has to match the testing data for optimal performance. Consequently, training data created with measured textures as well as physically accurate material models and lighting should facilitate the best performance under real-world conditions. However, they conclude that more realistic lighting and reflectance models are not needed for training, based on the test performance achieved on the synthetic Sintel MPI Dataset [16]. A statement which, to my perception, contradicts the presented results and can only hold true for achieving optimal test performance on synthetic datasets alone.

3.4. Augmentation of Real-World Datasets

Data augmentation is commonly applied to training schedules and transforms the original data in 2D image space to improve network robustness and generalization [42]. The transformed images complement the original training set, assuming that the network perceives the transformed data as new. The type and severity of the augmentations have to chosen with great care as they directly influence the network performance with respect to critical appearance effects and application domains. *Phenomenological augmentation models*, including spatial transformations, chromatic transformations, noise and blur models, are often used due to computationally efficiency and the ability to increase robustness towards critical appearance effects. They are suited to increase the performance for classification and object detection tasks [141, 56, 146] and dense correspondences problems [94, 95]. Spatial image transformations apply cropping, flipping, and 2D affine transformations on the original image data. They are an effective way of ensuring that networks never see two identical images and, when applicable, beneficial to reduce over-fitting and increase generalization capabilities. Artificial noise and blur models, *e.g.*, Gaussian noise and blur, can be injected into the training data to mimic challenging camera aberrations such as severe sensor noise and lens or motion blur. The models typically lack physical realism, *i.e.*, they are applied

in a post-process on the tone-mapped intensity data and often ignore color reconstruction as well as scene context, but can still lead to improved performance and robustness for noisy and blurry testing data. Chromatic image transformations change the brightness, colors, and contrast of the image data and aim to mimic changes in the sensor setup and lighting. Just as the noise and blur models, the transformations are simple, applied in the image color space, and therefore physically implausible. Nevertheless, the augmentations are useful to increase robustness towards challenging lighting scenarios, *e.g.*, low contrast night and foggy scenes or saturated image regions, and invariance to imager color responses and reconstruction. The study in [56] extends the chromatic transformations with phenomenological models for snow, fog and a frosted camera lens for testing image classification networks. The models are simple overlays on the original image data that lack awareness of scene context. Therefore, the authors propose to not use the models for training but to assess robustness towards obstructions during network testing.

By contrast, more sophisticated augmentation techniques do not only rely on the image data, but also consider scene context. Alhaija *et al.* [1] present a method that can be used to augment real-world images taken from KITTI [38] with synthetically rendered cars. The cars are rendered with a ray-tracer and placed on top of the real-world images using reflection information from the supplied environment maps. In a post-processing step, the lighting of the real images is adjusted adjacent to the cars creating shadowing effects. This data can then be used to train neural networks for object detection of cars. They show that object detection networks perform better on real-world test data when they are trained on the augmented images opposed to the purely synthetic data of Virtual KITTI [35], The approach is extended in a follow-up paper [3]. AbuAlhaija2018 Here, real-world images are again augmented with synthetically rendered cars. Only this time, a DNN was used for rendering, exploiting semantic and geometric scene information. The baseline rendering quality of the DNN almost matches a physically based renderer. However, in comparison to the previous augmentation method, the synthetic cars better blend into the environment, as the approach can estimating realistic lighting interactions between the cars and real-world background. As a result, the test

performance of the trained object detector can be further improved by a small margin.

3.5. Summary

In this chapter, I presented various existing datasets for training, testing, and performance analysis of low-level correspondence problems and high-level vision tasks. I mainly assessed the rendering realism, *i.e.*, the light transport simulation and the material models and textures. To summarize my findings:

Initial low-level vision datasets are limited to simulating purely diffuse appearance through biased rasterization and Lambertian materials. Lately, there has been a shift towards using more sophisticated ray tracing to simulate specular and global lighting effects. However, only the MPI Sintel [16] and Flying Things[94] datasets are rendered with unbiased algorithms. The presented datasets utilize simplified material models and in no cases, measured reflectance. Instead, the textures are either artistically driven or created from a single photo. Consequently, the realism of each presented dataset is limited in one way or the other: Non-diffuse reflectance effects can either not be simulated at all, are simulated with physically implausible severity, or simulated with unrealistic spatial distributions that do not correspond to real-world materials.

High-level vision datasets focus on creating volume data, as most algorithms are neural network based. Consequently, biased real-time rendering systems with rasterization and screen-space global lighting approximations are most prevalent. While an accurate physically based material model is often used, the input textures for models are, again, artistically driven or derived from single photographs. One exception is the fully procedural high-fidelity dataset presented in [144, 161]. The images are rendered with unbiased path tracing, physically based material models, and material textures derived from photogrammetric measurements. Compared to rasterized datasets with lower realism, the training and testing performance can be vastly increased and almost matches real-world dataset performance.

I further presented studies on the differences between real and synthetic datasets in the context of computer vision testing, training, and performance analysis. For performance analysis and testing, the studies showed that even with realistic mate-

rial models and ray tracing systematic differences to real-world datasets remain [101]. Concerning training data generation, Mayer *et al.* [95] found that ideally, the rendering and material models of the training data should closely match the testing data. Additionally, they identified material textures as a decisive factor for training data quality. These findings indicate that for best performance in real-world applications, the rendering system should be accurate and the material textures of high realism.

My work in Chapter 4 addresses the systematic differences to real-world data. I show that BTF reflectance data from accurate and dense measurements can be used to create synthetic datasets with high perceptional realism and closely matching optical flow performance.

In Chapter 5, I present a method to fit BRDF models to such measured BTF data. I further study various material samples and assess the most critical reflectance effects for optical flow algorithms. I show that for all material types, an efficient fitted BRDF model can be used as a substitute for BTFs with virtually identical performance.

My work of Chapter 6, addresses training data augmentation to increase network robustness. Existing approaches mostly use phenomenological models to augment the original data in image space. The studies of Abu Alhaija *et al.* [1, 3] show that by considering scene context, real-world datasets can be realistically augmented with new synthetic elements. My presented method is closely related to those studies. However, it differs in the way that no new rigid objects are inserted, but the real-world appearance is modified with physically plausible precipitation effects.

4

Reflectance Field Renderings For Synthetic Optical Flow Datasets

This chapter is based on my work previously published in [47]

4.1. Motivation

CAN we render surfaces accurately enough such that optical flow algorithms cannot tell the difference to real-world surfaces? Are existing synthetic datasets accurate enough? These are the driving questions that motivated this work. Based on the findings of Vaudrey *et al.* [147] and Meister *et al.* [101], the answer is no. They state that synthetic data generally lacks critical effects, that affect the outcome of optical flow algorithms. Synthetic data is less discriminating for optical flow estimation and, consequently, the resulting flow fields are smoother. Therefore, they conclude that synthetic data cannot be applied with high confidence in testing for real-world tasks, as the lighting and material models generally lack physical realism.

Nevertheless, synthetic data has been successfully applied in training Deep Neural Network networks for optical flow estimation [26, 64]. Recently, material textures, as well as lighting and material models, have been identified as relevant factors for training data quality [95]. They show that realistic textures with large variety are essential for training data quality and that the lighting and material models of the training data have to match the testing data for optimal performance. Consequently, physical realism of the textures and physically accurate lighting and reflectance models should improve performance on real-world flow estimation. Unfortunately, they conclude that more realistic lighting and reflectance models are not necessarily needed for training, based on the test performance achieved on the synthetic Sintel MPI Dataset [16].

However, the current approaches that are commonly used for training and testing optical flow display several shortcomings for texturing and modeling surface reflectance and the light transport simulation. Most datasets are created with biased rendering techniques, *i.e.*, mostly rasterization, and simple material models, *i.e.*, purely Lambertian BRDFs or simple Phong BRDFs. The reflectance textures are artistically driven and do not include spatially varying specular terms (*c.f.*, Chapter 3). As a result, the textures are often physically implausible and too crisp, the specular highlights are faint and lack spatial variations, Fresnel reflections are missing and complex light paths, *e.g.*, specular interreflections and caustics, cannot be simulated.

On the other hand, computer graphics provide the means for more accurate simulations. Physically based light transport algorithms can be used to remove bias from the datasets (*c.f.*, Section 2.3). Sophisticated measurement devices can accurately capture surface reflectance as an equivalent of the 9-dimensional BSSRDF function (*c.f.*, Section 2.4).

In the following sections, I present my findings of exploring the upper limits of computer graphics, *i.e.*, state-of-the-art texturing methods and reflectance models with unbiased light transport, in the context of optical flow performance analysis. The work presented here is solely my contribution unless explicitly stated otherwise and included for the sake of readability.

4.2. Contributions

This work explores the usability of high-fidelity synthetic data for optical flow performance analysis. I analyze the limits of synthetic data that is rendered with the bidirectional texture function (BTF) reflectance data, which was acquired with a state-of-art surface reflectance measurement device by the Christoph Schwartz of the University of Bonn.

The state-of-the-art measurement devices [131] for surface reflectance utilized a set of more than 50000 high dynamic range images to capture the reflectance fields of objects. The sampled reflectance fields are then stored in a BTF employing spatial and angular resampling and SVD compression. On first sight, unbiased path-traced renderings of the BTF are almost indistinguishable from real sequences. Only minor differences can be observed for highly reflective surfaces with high-frequency spatial

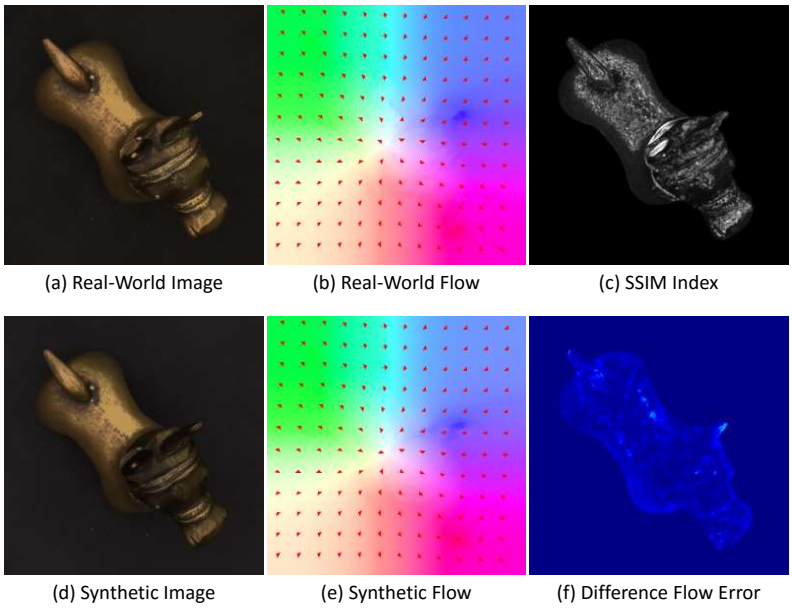

Figure 4.1.: A sophisticated dome setup [131] was used to capture the reflectance fields of two objects, a highly reflective donkey statue, and a mostly diffuse pyramid. Four stop motion sequences were shot and synthetically replicated, creating real-world images (a) and supposedly identically looking synthetic images (d). The structural differences between the images were assessed with the SSIM Index [169] (c), locating perceptually different structures (marked as white). The images sequences then served as input for a given optical flow algorithm. The computed flow fields of the real sequences (b) and the computed flow fields of the rendered sequence (e) are then compared, resulting in a difference flow error image (f). The lower the difference, the more appropriate the rendering method is for optical flow evaluation. The highest differences occur around delicate structures with high specular reflectivity and moving highlights, such as the donkey's ear tip.

variations in specular detail. But how do optical flow algorithms perform on such renderings? How do the simple renderings of most existing datasets compare? To answer these questions we followed the workflow depicted in Figure 4.1. I compare the BTF reflectance model to a simple Lambertian BRDF texture (TEX), a model that is representative for many existing optical flow datasets and least discriminating to optical flow algorithm in terms of simulated reflectance effects. I conduct both a perceptual analysis of the structural similarities between the synthetic and real-world sequences as well as an analysis of differences in computed flow fields.

Four real-world sequences of two objects with varying degree of complexity were created and synthetically replicated with both reflectance models. An unbiased physically based renderer was used to ensure that no additional bias from the renderer is limiting the synthetic data quality, possibly creating differences in the computed flow fields. Unbiased rendering algorithms are, to my knowledge, not used in any published dataset.

I then assessed the perceptual quality of the sequences in a quantative manner using the Structural Similarity Index (SSIM) [169]. The SSIM measures the structural similarity between two supposedly identical images. Structural similarity is a relevant aspect for optical flow simulation quality, as the motion of structures induces motion in the flow fields. Other standard metrics, such as the mean square error and the peak signal-to-noise ra-

tion, primarily assess differences on the pixel and global level, *i.e.*, noise and global lighting changes. Optical flow algorithms are, however, robust to such image perturbations.

The perceptual analysis reveals: BTF rendering simulate optical flow corner cases accurately, including specular highlights, subsurface scattering, and delicate geometric details. However, the accuracy is limited by the sampling density of the capturing devices. Increased geometric, angular and spatial resolutions, as well as less aggressive SVD compression and BTF resampling, can lead to increased visual quality. At this moment, the angular sampling rate is the most limiting factor for the most complex objects with mirror-like reflections and delicate geometric details. A high geometric resolution is beneficial for objects with small geometric detail, such as cavities.

The then conducted optical flow experiments confirm this assessment, utilizing a difference flow error metric (DFE) to assess the magnitudes of the differences in the flow fields. BTF renderings are, in general, a sufficiently accurate reflectance model for optical flow performance analysis. They can handle materials with subsurface scattering and medium to strong specular highlights. The errors in the flow fields are significantly reduced compared to simple TEX renderings and most often well within the sub-pixel accuracy of the used algorithms. More substantial errors remain when mirror-like reflections and small geometric details are combined. However, the errors are significantly reduced in comparison to simple renderings. The BTF results lead to the presumption that the error could be removed entirely with an increased sampling rate for arbitrary opaque material types.

When limiting oneself to less challenging materials without mirror-like reflections and small geometric detail, the reflectance field renderings are indistinguishable to real-world images for optical flow algorithms. Simple renderings, on the other hand, are not sufficiently discriminating for optical flow algorithms and, in general, only sufficient for Lambertian surfaces without subsurface scattering and small geometric details.

4.3. Method

In this section, I describe the general setup of our experiments, the generation of the test sequences and setup for the optical flow performance analysis. The planning of the test sequences was

conducted jointly with the University of Bonn. Measurements have been solely carried out by the University of Bonn with an existing measurement setup [131]. The acquisition and creation of test scenes are described in detail to provide the necessary context. I frequently refer to the corresponding publications for the measurement procedure and setup. I conducted the subsequent optical flow experiments and the following evaluations.

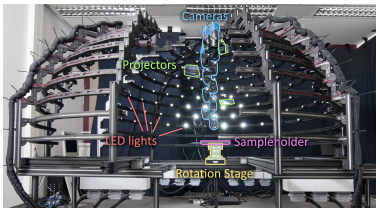

Figure 4.2.: The Dome II BTF acquisition device at the University of Bonn. The setup, which is comprised of a hemispherical structure, is slid open to give view on the components on the inside.

4.3.1. Acquisition of Geometry and Surface Reflectance

We used the Dome II BTF acquisition setup [131] of the University of Bonn, depicted in Figure 4.2. It captures the 3-dimensional geometry and 9-dimensional reflectance fields of objects, that are placed in the center of the dome. The overall object digitization procedure follows the approach proposed by Schwartz *et al.* [130]. First, structured light from the digital projectors is used to triangulate the object in different rotation stage poses. The resulting point clouds are then registered, and the surface is reconstructed to form a single 3D mesh. The object is rotated on the turntable in 15° steps, to capture anisotropic the reflectance. For each rotation, the 198 light sources are lit sequentially, and the resulting appearance is captured with 11 cameras. The cameras use multiple exposures to capture the linear reflectance in high dynamic range accurately. Finally, the resulting images are projected onto the object surface parametrization, the directions are resampled into local coordinate frames, and the data is compressed using SVD based dimensionality reduction. Details on this procedure can be found in the referenced literature [130] and [131].

The objects in our test scenes have been separately digitized. The 3D geometry is stored as a triangle mesh. We chose the number and distribution of triangles to sufficiently preserve the small geometric details. In the extreme case of the blackened plates, which was completely flat, this resulted in a single quadrilateral. The captured BTF reflectance is given as tabulated values at 151×151 discrete view and light directions, both equally distributed on the hemisphere, and a number of spatial positions on the surface, densely covering the whole object. See Table 4.1 for additional details. The resulting geometry and BTF reflectance data was then used as an input for the state-of-the-art, open source, physically based renderer “Mitsuba” by Jakob

Table 4.1.: Attributes of the captured objects. The column "# texels" denotes the number of data points on the object's surface for which the bi-directional reflectance is stored. To evaluate the influence of the spatial resolution, we created two different BTFs from the donkey data.

Object	**Geometry**		**BTF**	
	# Δ	avg. edge	# dir's	# texels
Donkey 3k	71k	0.94 mm	151^2	3072×3072
Donkey 2k	71k	0.94 mm	151^2	2048×2048
Pyramid	75k	0.66 mm	151^2	1600×1600
Plate 1	2	275 mm	151^2	512×512
Plate 2	2	203 mm	151^2	512×512

Wenzel[1].

4.3.2. Generation of Test Scenes

To evaluate optical flow performance, two separate test scenes have been constructed, *c.f.*, Figure 4.3. The first scene consists of a brass donkey statue positioned on a rotating blackened plate, the second one of a clay pyramid on a slightly smaller plate. The surface of the metallic donkey statue boasts significant specularity. The surface of the dielectric clay pyramid contains only minimal specularities and some more notable subsurface scattering.

For both test scenes, two real-world sequences with distinct camera angles have been shot under controlled illumination conditions. Here we used the Dome II measurement device as well, which allowed us to have accurate geometric calibration data and known radiometric camera and light characteristics. The exact calibration procedure is described in [131] and yields average reprojection errors of 0.16 pixels for the cameras and an error of about 0.08° for the light directions. Each sequence consists of 11 frames recorded in stop-motion with the plate rotating one degree between each frame. To achieve the impression of ambient illumination, we activated all 198 LED light sources simultaneously.

Rendering with reflectance from BTF data was implemented as a new material plug-in to Mitsuba. Because of the BTF's far-field assumption, the appearance does not depend on the position that light enters the surface. Hence, reflectance can be considered a local phenomenon and a standard path tracing integrator can be used. During rendering, our plug-in uses linear interpolation between the discretized data points to compute the radiance for a given sample (position and incoming and outgoing direction). The data points are reconstructed on-the-fly from the

[1] http://www.mitsuba-renderer.org/

compressed SVD representation.

For rendering with 2D textures, the diffuse material plug-in that ships with Mitsuba has been employed. All settings have been chosen similar to the BTF rendering. The textures for the objects have been generated from the BTF measurement data. Here, we loosely followed best practices for texturing 3D models from photographs: For each texel, we computed the radiance for the view-direction perpendicular to the surface under homogeneous ambient lighting conditions.

Virtual camera parameters, light source positions, and light source intensities were chosen to match those of the real test sequences, for which the accurate Dome II calibration data was available. Note that the illumination and the turntable rotations used in the test sequences were not part of the input data to generate the BTFs but are novel conditions. Furthermore, as the objects have been captured separately, the indirect lighting between the blackened ground plate and the donkey or pyramid in the synthetic scene was also not part of the captured reflectance but was simulated by the path tracer as well.

In both cases, we used 128 samples per pixel to avoid rendering noise. The images were then super-sampled with twice the resolution in both directions, effectively leading to 512 samples per pixel for the final input images of the optical flow algorithms Otherwise, the flickering appearance between consecutive frames might have influenced the results of the optical flow.

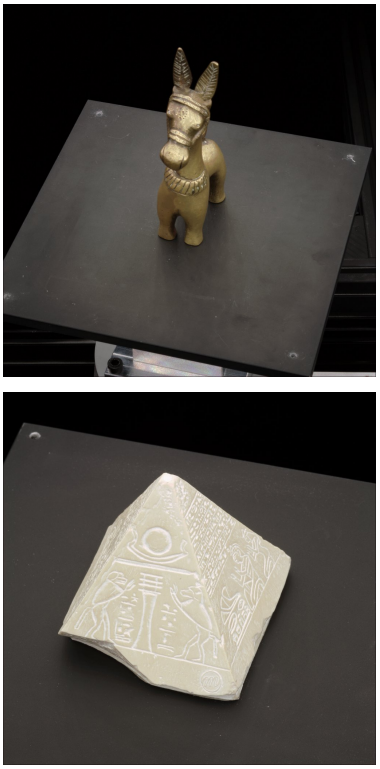

Figure 4.3.: Real photographs of the two scenes used in our experiments. In the synthetic renderings only the object and the blackened ground plate are considered.

4.3.3. Experimental Setup

The method above describes the acquisition and generation of test sequences that were planned with and provided by the University of Bonn. Once provided, I performed a perceptual analysis and optical flow experiments on the sequences.

The goal of the quantative perceptual analysis was the assessment of the visual appearance and the structural similarities between the synthetic and real-world sequences. With this assessment, I could better identify the critical reflectance effects for optical flow performance analysis and the limitations of the used reflectance models. The visual differences can be assessed with the mean square error metric or the peak signal-to-noise ratio, but they mostly focus on image noise and global lighting differences. Optical flow, however, is most sensitive to the motion of image structures. Hence, I utilized the structural similarity

index as a performance metric.

For the optical flow experiments, I computed the optical flow on each image pair of the sequences using three different methods: A multi-scale, nonlinear version of the Horn&Schunk method [62] similar to the one described in [100]. Secondly, the method of Papenberg *et al.* [116] was used with Charbonnier-penalizers both applied to the regularizer and the intensity-only data term. Both methods were implemented with the open source Charon Framework [43] using five pyramid levels and a downsampling factor of two. As a third method the publicly available MDP-Flow2 Algorithm by Xu et al. [163] was used. It is one of the best-performing methods on the Middlebury website and comes with a freely available implementation. Parameters for all methods were chosen to produce flow fields as correct as possible. The differences between the synthetic and real-world flow fields were then analyzed through our own difference flow error metric. This metric was used to assess the flow differences both globally and locally on the areas with critical reflectance effects.

Structural Similarity Index

In the context of assessing the image quality of our simulations, we are specifically interested in the differences between the real and synthetic images. The mean square error and peak signal-to-noise ratio can be used to assess the absolute differences between images. Both are purely intensity based measures and well suited to asses global intensity changes and image noise. Optical flow performance, however, mostly depends on distinct structures and to a much lesser extent on noise or global intensity changes. Structures are necessary to find correspondences between image pairs. In uniform regions, optical flow algorithms rarely perform well. Changes in the structural appearance, such as specular highlights or blurred edges, can therefore negatively affect the performance of optical flow algorithms. Therefore, those effects have to be simulated accurately in a discriminating synthetic dataset. Noise, on the other hand, can negatively influence the performance of optical flow algorithms but is mostly compensated by regularization of advanced algorithms. Global lighting changes cause an offset in the data term of the flow problem, which can generally be handled by the least square optimization.

The structural similarities can be assessed with the structural

similarity index (SSIM) [169]. It is a well-established measure that compares the contrast, luminance, and structures of two supposedly identical images. A $SSIM = 1.0$ means the images are identical and a $SSIM = 0.0$ means that the images do not resemble each other in any way. An SSIM index map can be visualized, where each pixel is encoded with its corresponding SSIM contribution. I used an inverse color coding for the visualization, to provide easier comprehension of the printed figures. An SSIM index map should therefore be white for an image with $SSIM = 0.0$ and black for an image with $SSIM = 1.0$, respectively.

It should be noted, that while an SSIM of close to one should be the goal of accurate simulations, it is in practice not achievable with our experimental setup. There are several parameters of our virtual scenes, beyond the choice of the reflectance model, that affect the simulation quality of the sequences. The virtual scenes are carefully designed replicas of the real-world conditions and parameterized with calibrated dome data. However, the calibration data is inherently afflicted with measurements errors, which translate to errors in the virtual scenes. In the following I list the most relevant aspects and their corresponding errors induced in the synthetic images:

- *Geometric light calibration accuracy:* The light sources are positioned on the hemisphere of the dome setup and then carefully calibrated. Schwartz *et al.* [132] claim an accuracy of 0.08° for the measured light source positions, which corresponds to a reprojection error of roughly $0.4px$. The virtual light sources are placed at the calibrated positions, and the reprojection error is induced in the synthetic sequences.

- *Radiometric light calibration accuracy:* Two factors cause different power distributions for each light source: i) The LED bulbs have a production tolerance for the spectral power distribution of the light and thus different brightness levels. ii) The light is focused with a manually aligned lens-optic onto the material sample. The radiometric calibration corrects both factors. However, the radiometric calibration assumes a flat material sample and is therefore erroneous for protruding objects [132]. Unfortunately, the resulting error is unknown. The radiometric calibration

was applied to the virtual light sources, which are consequently affected by this unknown error.

- *Geometric camera calibration accuracy:* The cameras are fixed at 11 distinct positions in the used dome setup. However, the measured object is rotated with limited accuracy. According to Schwartz *et al.* [132], the errors of the camera calibration are a spatial error of $11\mu m$ and an angular error of $0.001°$, corresponding to a reprojection error of roughly $0.16px$. The virtual camera extrinsics are chosen to accord to the calibrated extrinsics, and the reprojection error is induced in the synthetic sequences.

- *The camera noise:* The cameras used in our experiments are the VS Vistek SVCam CF 4022COGE, which is an industrial grade camera with an average signal-to-noise ratio of 32dB [132]. The noise corresponds to approximately 0.16 pixel intensities in tone-mapped 8-bit color space and is not simulated in our synthetic sequences.

In addition to the errors induced by the reconstruction of the virtual scenes, the path tracing integrator afflicts the synthetic images with *rendering noise* of an unknown magnitude. In the worst case, this error is 0.5 pixel intensities in tone-mapped 8-bit images, but in reality, it should be notably smaller as all light sources are directly sampled on each light surface interaction.

While all mentioned errors certainly affect the synthetic images in a way, only the *geometric light calibration accuracy* and *geometric camera calibration accuracy* should affect the structural similarities of the images in a meaningful way. The *rendering noise* and *camera noise* are small and the *radiometric light calibration accuracy* induces small frequency lighting errors, which hardly affect the structural similarities.

Difference Flow Error Metric

In the context of synthetic datasets for optical flow performance analysis, we are specifically interested in differences between the flow fields computed on the real and synthetic images. Following the definition of the endpoint error [114], which is commonly used in optical flow evaluation, the difference flow error (DFE) can be defined as the magnitude of the difference between the flow fields computed on the synthetic and the real-world data:

$$DFE = \sqrt{(u_{real} - u_{synth})^2 + (v_{real} - v_{synth})^2} \tag{4.1}$$

The DFE was analyzed for all flow fields computed on synthetic sequences. It was evaluated only where the actual content is visible, while the mostly black background did not contribute to the error.

It should be noted, that while a DFE of close to zero is the goal of synthetic sequences, it is in practice hardly achievable in this experimental setup. While reflectance models certainly account for the most substantial systematical errors in the flow fields, as shown throughout this analysis, some remaining differences between the flow fields of synthetic and real-world datasets have to be expected. Two aspects limit the DFE: i) The *virtual scene accuracy* that leads to differences in the visual appearance, *i.e.*, errors induced by the extrinsic calibration of the light and camera positions, the intrinsic camera noise, the radiometric calibration of the light sources and the noise of the rendering method. For a detailed explanation, I refer to the previous section 4.4.1. ii) The *accuracy of the optical flow algorithm* that is inherently limited. Advanced flow algorithms have an accuracy of less than $0.2px$ for carefully engineered scenes of diffuse objects with small object and camera motion (*c.f.*, Middlebury dataset benchmark results as of the publication of this thesis.[2] Those kinds of scenes are the least discriminating to optical flow algorithms and their underlying diffuse world assumption and thus constitute optimal flow conditions and best possible performance.

In the following, the flow results and the DFE are analyzed with respect to the hypothesis that synthetic datasets can achieve suitable realism for optical flow. To this end, the DFE is divided into magnitude-related categories:

- $DFE \leq 0.05px$: The DFE indicates that the pixels belong to regions that are *virtually identical* in synthetic and real-world flow fields.

- $DFE \in [0.05, 0.2]px$: The DFE indicates that the pixels belong to regions whose accuracy falls within the *accuracy of the optical flow algorithm*. The regions are faithfully

[2] http://vision.middlebury.edu/flow/eval/results/results-e1.php

replicated, and errors from the reflectance model can be mostly precluded.

- $DFE \in [0.2, 0.5]px$: The DFE indicates that the pixels belong to regions whose accuracy falls within the *accuracy of the optical flow algorithm* and the *virtual scene accuracy*. The regions are still faithfully replicated, but smaller errors of the reflectance model cannot be precluded.

- $DFE \in [0.5, 1.0]px$: The DFE indicates that the pixels belong to regions where *medium errors* are likely elicited from the reflectance model. The regions are not replicated faithfully, which could be caused by errors of the reflectance model in addition to the errors induced by the virtual scene reconstruction.

- $DFE \geq 1.0px$: The DFE indicates that the pixels belong to regions where *large errors* are likely elicited from the reflectance model. The regions are not replicated faithfully, and the errors cannot be explained by reconstruction errors alone, leading most certainly to errors from the reflectance model.

It should be noted, that those DFE ranges are empirical and not meant as absolute thresholds. They act as a guideline to better understand and assess the DFE results in the following analysis.

The first range of $[0.05, 0.2]$ was roughly correlated to publicly available benchmark endpoint error results for the real-world sequences of the Middlebury dataset [6]. The other ranges were derived from the least challenging sequence, the top view sequence of the mostly diffuse pyramid. The $[0.2, 0.5]$ error range roughly correlates to the DFE values observed in the non-damaged regions of this sequence. Those regions display perfect conditions for optical flow algorithms. They are structured but almost perfectly diffuse with hardly any motion dependent appearance. The $[0.5, 1.0]$ error range roughly correlates to the DFE values observed in the more challenging parts of this sequence, the fissures. In those regions the appearance of the surface is slightly motion dependent, inducing small additional errors in the flow fields.

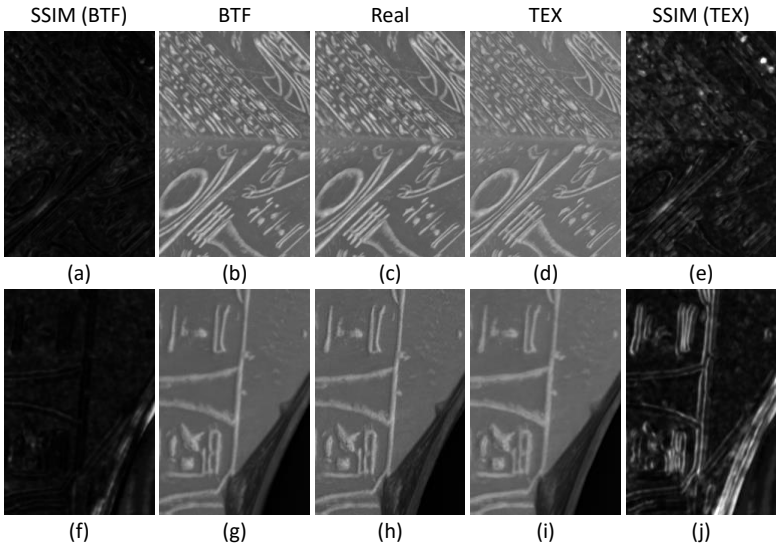

Figure 4.4.: The images depict close-up images of the pyramid sequences and the SSIM error maps. The real-world images (c) and (h) exhibit mostly diffuse reflectance and some minor shadowing and subsurface scattering at the fissures in (c) and (h) and the chipped of parts in (h). The corresponding TEX renderings (d) and (i) are overall brighter than the real-world images. Global lighting differences are, however, unproblematic for optical flow algorithms. Medium structural differences can be observed at fissures and at the bottom area in (i), which faces away from the acquisition setup, *i.e.*, light and cameras. This is visualized in the inverse SSIM maps in (e) and (j). The BTF renderings in (b) and (g) do not exhibit such a global lighting change and can almost perfectly simulate the reflectance effects of the real-world images. This is reflected in the SSIM maps (a) and (f) by low error values. The largest remaining differences are located at the chipped off bottom area in (f), which is facing away from the acquisition setup.

4.4. Results

In this section, I describe the experiments conducted on the provided image sequences. I start with the perceptual analysis that was conducted to get the first indication of possible BTF and TEX model inaccuracies. Then I continue to the optical flow experiments performed on the image sequences. I show, to what extent the small inaccuracies in the BTF renderings and the large inaccuracies of the TEX renderings translate to a diverging optical flow performance.

4.4.1. Perceptual Analysis

In this section, I present the results of the perceptual analysis performed on the real-world images and the renderings with two types of reflectance models, *i.e.*, the sampled reflectance fields (BTF) and the Lambertian BRDF Textures (TEX), for two mostly diffuse pyramid sequences and two more challenging donkey sequences. I focus on the structural differences between the synthetic and real-world images of the pyramid and donkey sequences, utilizing the SSIM metric. I then continue to general observations on limitations of the BTF reflectance model and potential improvements.

Pyramid Sequences The clay surface material of the pyramid is dielectric and can be considered almost Lambertian with minimal specular reflectivity. It boasts some shadowing and subsur-

face scattering, most notable in the damaged regions and fissures on the surface. Example close-up images of the pyramid top and frontal view sequences and the corresponding renderings with their structural similarities compared to the real-world images are depicted in Figure 4.4.

Overall, the renderings of the TEX reflectance are roughly matching the appearance of the real-world images, with structural similarities of 0.825 in the top view crop and 0.77 in the bottom view crop. However, the renderings are slightly brighter than the real-world images. Likely, this is because the diffuse texture taken from the top view and camera positions includes specular lighting information of the pyramid. This is not necessarily problematic for optical flow algorithms, as they are mostly invariant to global brightness changes. The shadowing and subsurface scattering at the fissures cannot be replicated correctly by the TEX model. This translates to notable errors in the SSIM index map. The most substantial errors are visible at the bottom of the pyramid. At this area, the surface is facing away from the structured light scanners as well as all cameras and light sources. As a result, the geometry and the diffuse texture could not be reconstructed accurately.

The renderings of the BTF reflectance model are more closely matching the appearance of the real-world images, with structural similarities of 0.887 in the top view crop and 0.869 in the bottom view crop. In contrast to the TEX model, no global brightness increase can be observed. The shadowing and subsurface scattering at the fissures is replicated correctly, and only minor errors are visible in the SSIM index map. The errors for the critical bottom area are significantly lower using the BTF model compared to the TEX model. The BTF reflectance model is capable of mostly compensating the inaccuracies of the geometry scan in this area, due to the reflectance field assumption. The remaining differences occur due to an inaccurate silhouette and the generally low sample count of light sources and cameras in this area.

Donkey Sequences The brass donkey statue exhibits medium to strong specular reflectivity combined with smooth areas at the torso and delicate geometric detail around the head area. Example images of the donkey head and torso areas and the corresponding renderings with their structural similarities compared to the real-world images are depicted in Figure 4.5. Spec-

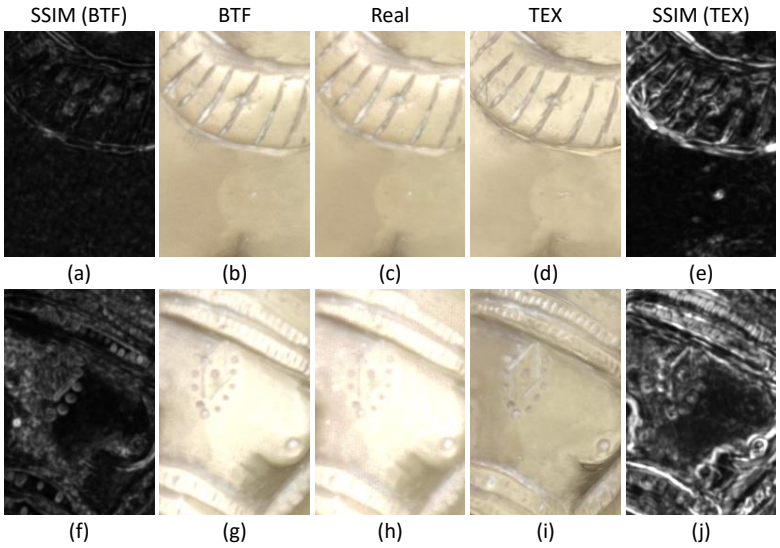

Figure 4.5.: The images depict donkey close-up images and their SSIM error maps. inter-reflections. The real-world image in (c) contains specular highlights of low intensity at the bottom flat torso area and of medium intensity at the geometric structures at the top. The image in (h) depicts the head area with mirror-like reflections and delicate geometric detail. Both real-world images have blurred cavities due to specular inter-reflections and Fresnel reflectivity. As expected, specular effects missing in the TEX renderings in (d) and (i) and contours are significantly more pronounced, favorable conditions for optical flow. The SSIM errors are large in those areas in (e) and most notably in (j). The BTF renderings in (b) and (g) are significantly more accurate, simulating both highlights and blurred cavities. However, minor errors remain at the contours, which is visualized in the SSIM map of (a). I attribute this to the low sampling rate at low grazing angles, where the Fresnel reflectivity increases rapidly. Slightly larger errors occur around the highly reflective head area depicted in (f). The angular sampling limits the accuracy for areas with mirror-like reflections and delicate geometric details and therefore. However, the simulation has significantly improved compared to the TEX model.

ular highlights can be observed in the real-world images where the surface normal points in a way that most light is reflected towards the viewer. Strong highlights are visible at the top area of the torso close-up and the polished area of the head close-up. Notice, that there are small spatial variations in the specular highlights. The cavities on the object's surface appear blurry due to Fresnel reflectivity and light inter-reflections, *i.e.*, light bounces multiple times within the cavities before it gets reflected towards the viewer. Both effects are critical for optical flow applications: Algorithms can falsely detect motion from highlights that are moving along with the light source and algorithms can lose correspondences when structures are blurred. As expected, the TEX model cannot recreate those effects, and the structural similarity is low, *i.e.*, an SSIM of 0.850 for the torso rendering and 0.672 for the more complex head rendering. The specular highlights are naturally missing, and the appearance is homogeneous in the flat surface areas. Cavities are clearly visible and significantly more pronounced than in the real-world reference. Those are easy conditions for optical flow algorithms. Critical specular effects do not induce false motion, and clear pronounced structures simplify the correspondence estimation. The BTF reflectance model is better suited to replicate such a specular object. The structural similarities are higher with an SSIM of 0.923 for the torso rendering and 0.859 for the more complex head rendering. The smooth areas with medium reflec-

tivity of the torso are rendered without any noticeable visual differences. In comparison to the TEX model, the simulation of the cavity blurring has significantly improved. However, it is not simulated to the full extent, and small differences remain at the top area of the torso rendering and the entire head area. Likely, this is because the BTF sampling does not include low grazing angles, causing Fresnel reflectivity to be inherently less accurate. In contrast to the TEX model, specular highlights can, in general, be simulated. However, considerable differences remain. The spatial variations of the highlights are missing in the top torso area, and mirror-like highlights of the head area are fainter.

General Observations Highly reflective areas with small delicate structures appear to be the main challenges for BTF reflectance field renderings. This can be explained:

Matusik *et al.* [93] observe that an angular resolution of approximately 1° is required to capture mirror-like reflections. The angular resolution of the dome setup is significantly lower with 7.6° for viewing and 9° for lighting directions. This sampling rate was chosen to best to fulfill the Nyquist criterion of common cultural heritage objects while limiting measurement time and data storage at the same time [132]. For flat homogeneous areas with similar surface normals, this can be compensated from neighboring pixels. More precisely, the likelihood that all pixels of the area miss the specular highlight on all view and light directions is low. The captured reflectance fields are resampled to the BTF format before rendering. The final BTFs are constructed for 151×151 view and lighting directions and a resolution between 1600×1600 and 3072×3072 (cf. table 4.1). This results in an angular resolution of 9.4° in viewing and lighting directions and a spatial resolution of $190DPI$. By this resampling, small reflectance details, such as variations in the specular highlights, can get blurred. This blurring is further amplified by the compression of the sampled BTF data to an SVD representation. The inaccuracies of the geometric sampling with respect to the surface reflectance are intercepted by the reflectance field approximation and therefore negligible [23]. However, coarse geometric sampling can affect the silhouettes of the object shape. Therefore, small geometric details below the average edge length of $0.94mm$, such as the observed cavities, are inherently defective.

Overall, the quality of the BTF reflectance model seems to be limited by the sampling density of the acquisition process. An increased angular resolution is beneficial for highly reflective materials. A denser spatial sampling of the reflectance field can decrease the likelihood that small specular highlights are missed. Forfeiting the resampling and SVD compression can be beneficial for spatial variations of the reflectivity. An increased geometric sampling can decrease the errors at cavities. However, all those measures further increase the data storage and the already demanding acquisition and rendering times.

4.4.2. Optical Flow Analysis

In this section I present the optical flow results on real-world images and renderings with two types of reflectance models, *i.e.*, the sampled reflectance fields (BTF) and the Lambertian BRDF Textures (TEX), for two mostly diffuse pyramid sequences and two more challenging donkey sequences. The flow fields were computed with three optical flow algorithms: The basic multi-scale nonlinear Horn&Schunk (H&S) algorithm and two more sophisticated methods with the Papenberg (PPB) and the MDP-Flow2 (MDP) algorithms (*c.f.*, Section 4.3). I focus on the differences between the flow fields of the synthetic and real-world images of the pyramid and donkey sequences, utilizing the presented DFE metric.

Pyramid Sequences

The pyramid sequences depict a rotating clay pyramid with mostly diffuse reflectance, observed from a *Top View* and *frontal view*. Specular effects are minimal, and there is little geometric detail apart from a few chipped off areas and fissures. They were chosen as exemplary sequences for many real-world materials and can be considered a relatively easy task for optical flow algorithms with their underlying diffuse world assumption.

Only minor differences in the appearance of the pyramid could be identified via the perceptual analysis of the renderings for the BTF and TEX models (*c.f.*, 4.4). Most structural differences occur at the damaged regions at the bottom and the fissures on the surface. The probable causes are shadowing and subsurface scattering effects that cannot be modeled by the Lambertian BRDF model. The TEX model exhibits a slightly brighter appearance overall, which is expected to be unproblematic for

Method		PPB	H&S	MDP
Pyramid Top View	BTF	0.153 ± 0.133	**0.124**$\pm$0.070	**0.164**$\pm$0.104
	TEX	**0.136**$\pm$0.114	0.128 ± 0.071	0.166 ± 0.105
Pyramid Frontal View	BTF	**0.158**$\pm$0.118	**0.145**$\pm$0.105	**0.194**$\pm$0.143
	TEX	0.181 ± 0.139	0.151 ± 0.101	0.197 ± 0.142

Table 4.2.: Mean difference flow errors and standard deviation for the donkey sequences with all optical flow algorithms and rendering techniques. The best rendering is marked in **bold**. The means are reduced with BTF model, except for the PPB algorithm of the *Top View*. However, the overall differences between the models are minimal and no significant outliers exist in any case, reflected by the low standard deviations.

optical flow algorithms In general, all synthetic sequences closely match the structural appearance of the corresponding real-world images.

This directly translates to computed flow fields. Table 4.2 depicts the mean DFE for the *Top View* and *frontal view* pyramid sequences and all combinations of synthetic sequences and optical flow algorithms. The displayed values were computed over all image pairs contained in each sequence. In the following, I present a detailed assessment of both pyramid sequences.

Top View: All three flow algorithms perform almost identical for the renderings of both reflectance models. The means and standard deviations of the DFE are slightly smaller using the BTF model for the H&S and MDP algorithms (*c.f.*, Table 4.2) and slightly larger for the PPB algorithm. Overall, the mean difference flow errors for this sequence are minimal and fall within the accuracy of the optical flow algorithms. The histogram in Figure 4.6 depicts the DFE distribution for the MDP algorithm. It shows that more than 70% of the pixels have a DFE below $0.2px$ and are well within the accuracy of the algorithm. Less than 1% of the pixels have a DFE larger $1.0px$ and can be considered outliers.

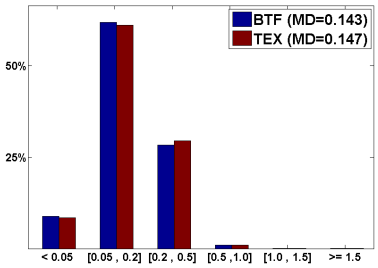

Figure 4.6.: Histogram of the difference flow error in the Top View Pyramid Sequence using the MDP-Flow2 algorithm. BTF and the TEX reflectance model yield almost identical flow fields, also reflected error distributions. Fewer than 1% of the pixel have a DFE larger than $1.0px$ while more than 70% of the pixels have very low sub-pixel error that fall well within the accuracy of the algorithm (DFE $\leq 0.2px$).

Frontal View: Compared to the *Top View* the means and standard deviations of the flow fields are slightly increased (*c.f.*, Table 4.2), but the results are fundamentally similar. In contrast to the *Top View*, the BTF reflectance model performs best for all three flow algorithms, and the margin between the reflectance models is generally more pronounced. In the DFE histogram of the MDP algorithm, we can observe that a few more large errors of $0.5px - 1.0px$ are now present in the flow fields. A visualization of the DFE for an exemplary image pair is depicted in Figure 4.8. The areas with the most prominent errors are the left surface, which faces away from the camera, and the larger damaged parts in the front. When using the BTF model these areas are smaller in size at the front surface and not as many

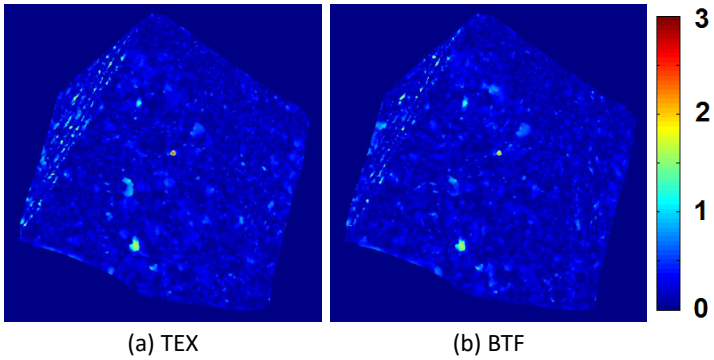

Figure 4.8.: Visualization of the difference flow error using the MDP-Flow2 algorithm on the pyramid Frontal View. Higher errors for the TEX model (a) occur around the fissures and chipped off areas on the front and the surface to the left, which is facing away from the camera. The underlying shadowing effects are not simulated. The BTF errors (b) are smaller in the front and fewer on the left surface, as shadowing effects are better simulated.

on the left surface, but not removed entirely. I attribute this error to the high-frequency shadowing and scattering occurring at the fissures. The artifacts at the fissures are similar for the PPB and H&S methods but less pronounced. The PPB method exhibits a few more large errors towards the object boundaries and the H&S method larger errors overall.

In both pyramid sequences, the errors are overall small for renderings with both reflectance models. Only a few medium outliers exist around fissures and at the damaged parts of the pyramid. This is where fine geometric detail creates shadowing and subsurface scattering effects. This indicates that a simple Lambertian BRDF model is a valid approach for synthetic optical flow datasets, as long as surfaces are smooth and almost perfectly diffuse. For mostly diffuse dielectric materials with fine geometric details, the BTF model can be used with almost perfect accuracy.

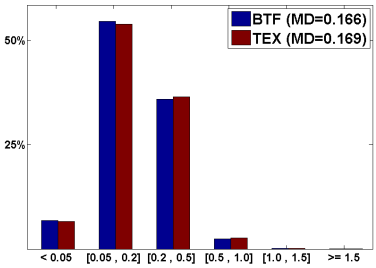

Figure 4.7.: Histogram of the difference flow error in the Frontal View Pyramid Sequence using the MDP-Flow2 algorithm. Compared to the Top View there are slightly more errors in the $[0.5, 1.0]$ range where the reflectance model might induce some flow errors. The BTF model performs slightly better in this range, but is similar overall.

Donkey Sequences

The second set of sequences shows a brass donkey statue on a rotating plate, observed from two distinct viewing positions, a *Top View* and a *frontal view*. The donkey contains regions with medium to mirror-like reflections and medium to high curvature. We have fine geometric detail and cavities around the head area of the donkey statue. This creates medium to large spatiotemporal changes of the appearance and a blurring of fine structures, a challenging task for optical flow algorithms. While the most challenging parts on the head of the statue are only partially included in the *Top View* sequence, all the mirror-like reflections, and small geometric details become apparent in the *frontal view*

Method		PPB	H&S	MDP
Donkey Top View	BTF2K	0.166 ± 0.135	**0.090**± 0.075	**0.177**± 0.195
	BTF3K	**0.165**± 0.125	0.102 ± 0.077	0.208 ± 0.212
	TEX2K	0.355 ± 0.390	0.178 ± 0.159	0.182 ± 0.196
	TEX3K	0.345 ± 0.397	0.169 ± 0.145	0.187 ± 0.194
Donkey Frontal View	BTF2K	0.249 ± 0.350	0.239 ± 0.317	0.378 ± 0.573
	BTF3K	**0.204**± 0.294	**0.223**± 0.311	**0.356**± 0.467
	TEX2K	0.364 ± 0.439	0.451 ± 0.689	0.411 ± 0.729
	TEX3K	0.341 ± 0.440	0.457 ± 0.705	0.406 ± 0.759

Table 4.3.: Mean difference flow error and standard deviation for the donkey sequences with all optical flow algorithms and rendering techniques. The best rendering is marked in **bold**. BTF renderings clearly outperform TEX renderings except for the MDP algorithm, where no HDR images could be used and the specular highlights are saturated. For the Top View sequence the BTF results can be considered almost as good as for the lambertian pyramid and fall well within the accuracy of the flow algorithms. In the more complex Frontal View the mean error is higher and the increased standard deviation indicates outliers. The TEX model performs poorly for both donkey sequences.

sequence.

Both effects can be simulated better with the BTF reflectance model (See Figure 4.5). However, the accuracy of the simulation is limited by the geometric, angular and spatial resolutions of the BTF. To better asses the limits of the BTF, two variants of the BTF model were created. A BTF model with a 2048×2048 texture (BTF2k) and a BTF model with a 3072×3072 texture (BTF3k). The angular and geometric resolutions are the more limiting factors in theory but are fixed in the used dome setup. Respectively, two variants of the TEX model were created with a 2048×2048 texture (TEX2k) and a 3072×3072 texture (TEX3k).

The specular effects of the statue can be better resolved in the 16-bit high dynamic range images provided by the University of Bonn. In the 8 bit images, the highlights often formulate larger regions with saturated uniform intensities, which is problematic for optical flow. Therefore, I conducted the donkey experiments on the original 16-bit data, whenever possible. The available implementation of the MDP algorithm can only handle 8-bit data. Consequently, the MDP algorithm performs poorly in those regions. The PPB results are more accurate and, therefore, presented in the detailed analysis.

In the following paragraphs I analyze how the observed visual differences translate to differences in the flow fields of the *Top View* and *frontal view* donkey sequences

Top View: The mean and standard deviations of the DFE for both reflectance models, both texture resolutions and all three optical flow algorithms are depicted in Table 4.3. The improvements of the BTF models contribute to a significant reduction of the mean DFE from $0.355px$ to $0.166px$ for the PPB algorithm and from $0.178px$ to $0.090px$ for the H&S algorithm. A noticeable error reduction cannot be observed with the MDP algorithm. As mentioned above, the 8-bit input images of the MDP

algorithm exhibit uniformly saturated regions, where specular highlights are present. Optical flow generally does not perform well on uniform regions, and the results are thus not representative.

Increasing the texture resolution, *i.e.*, using the 3K reflectance model variants does not noticeably reduce the errors for any combination of reflectance model and flow algorithm. As most visible parts of the statue exhibit only medium level specular reflectivity and little geometric detail, a vast improvement was not observed in the perceptual analysis and can therefore not be expected for the computed flow fields.

The large standard deviations for the TEX models and the PPB algorithm are striking. They indicate that some regions have significant DFE outliers. This cannot be observed for the BTF models, indicating that similar outliers are not present. Strikingly, large DFE standard deviations are not present for the H&S algorithm, and the DFE are overall smaller. This can be explained: The regularization of the H&S algorithm has to be generally stronger than for the PPB algorithm to perform adequately under arbitrary conditions. Consequently, the H&S flow fields are smoother than the PPB counterparts for both the synthetic and real-world sequences. Small details and structures are however missed. Therefore, the basic H&S algorithm is much less discriminating to simulation quality than the more sophisticated and accurate PPB. I chose the PPB algorithm for a more detailed assessment of the error distributions.

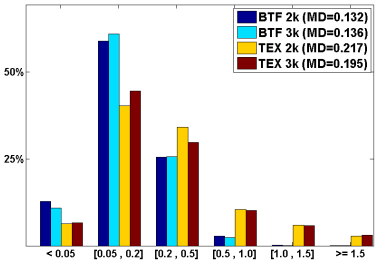

Figure 4.9.: The DFE histogram of the Donkey Top View Sequence using the PPB algorithm. Both variants of the TEX model perform poorly, with many outliers in the higher error ranges. The BTF models significantly reduce errors. Only a few errors remain in the $0.5, 1.0$ range, indicating errors induced from the reflectance model. The high-resolution BTFK3K variant performs slightly better than the BTF2K. Overall, the BTF distributions can be considered almost as good as in the pyramid sequences.

Figure 4.9 depicts the error distributions for the PPB algorithm on the *Top View* donkey sequence. For the TEX models, more than 20% of the DFE values are larger than $0.5px$ indicating that the reflectance model induces medium errors. Less than 50% of the DFE fall within the accuracy of the optical flow algorithms for the TEX2K model, *i.e.*, $DFE \in [0.05, 0.2]$, and roughly 51% for the higher-resolution TEX3K model. The rest of the DFE fall in the $[0.2, 0.5]$ range, where some small errors from the reflectance model cannot be precluded. The TEX3K has overall smaller errors, but the same amount of outliers. This indicates that substantial errors are tied to the reflectance model and cannot be offset by an increased spatial resolution.

The DFE can be significantly decreased using the BTF reflectance model. Less than 1% of the pixels have DFE values higher than $1px$ and only 3% of the pixels DFE in the $[0.5, 1.0]$

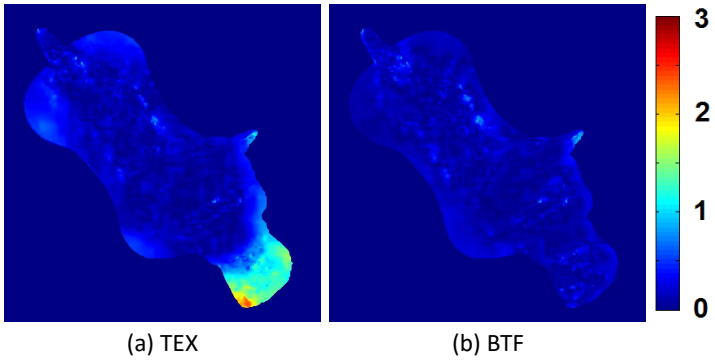

Figure 4.10.: DFE Visualization of the Donkey *Top View* sequence using the Papenberg algorithm. High DFEs occur around the head region in the TEX Rendering(b) due to insufficient simulation of specular effects. The BTF Rendering(a) reduced the DFE significantly. Medium errors remain where specular reflectivity is combined with delicate geometric detail, such as the approximately $1px$ errors at the ear tip to the right.

range. This totals to 4% of pixels that likely have flow field errors induced by the reflectance model. For both BTF models more than 70% of the pixels have low subpixel errors of less than $0.2px$ that fall within the accuracy of the flow algorithm and can, therefore, be considered perfectly accurate for optical flow performance analysis. The remaining 25% of the DFE fall in the $[0.2, 0.5]$ range, where some small errors could be induced from the reflectance model. Both BTF models perform similarly in terms of the mean DFE. However, the higher-resolution BTF3K model exhibits a generally narrower distribution with fewer outliers, indicating that critical parts are simulated a bit more accurately.

In Figure 4.10 you can see a Jet-Visualization of the DFE for an exemplary image pair of the donkey *Top View* sequence for the BTF3K and TEX3K models. This visualization helps to localize the improvements of the BTF model and the remaining differences. In the results of the TEX renderings, we observe high errors of up to $3px$ around the highly specular head region, errors of around $1px$ in back region and errors of around $1-1.5px$ at the ear of the statue. The largest errors in the head region and the errors in the back region could be resolved with the BTF model. The errors of around the ear remain. The perceptual analysis of the synthetic images concluded that renderings of the BTF models exhibit notable structural differences in areas of high specularity and delicate geometric detail. This translates to the DFE values around the donkey ear, where both aspects are present.

Overall, the BTF results on the donkey *Top View* can be considered almost as good as in the Lambertian pyramid case. There are only minor increases in the mean and median DFE

values and only a few more errors in the $[0.2, 0.5]$ range, where the reflectance model could cause differences in the flow fields. Very few large errors around $1.0px$ remain in areas of high specularity and delicate geometric detail. The donkey *Frontal View* focuses on precisely those problematic areas. The analysis of this sequence in the following paragraph gives a better understanding of the limitations of the BTF model.

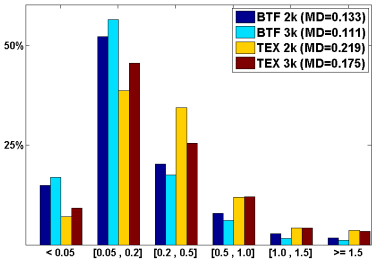

Figure 4.11.: The DFE histogram of our most complex sequence, the Donkey Frontal View, using the PPB algorithm. The TEX models perform similarly poorly as in the Top View. With the exception that the TEX3K model performs slightly better than TEX2K in this sequence. The BTF models significantly reduce errors. However, some larger errors remain for both variants. This indicates that the inaccurate simulation of the mirror-like highlights is affecting the flow performance. The BTF3K variant performs notably better than the lower resolution variant, especially in terms of larger outliers.

Frontal View: In this sequence, the donkey is observed from the front with a stronger emphasize on mirror-like reflections and the delicate structures of the head.

The mean and standard deviations of the DFE for both reflectance models, both texture resolutions and all three optical flow algorithms are depicted in Table 4.3. Compared to the *Top View* sequence the mean DFE is larger in all cases. A slight reduction of the DFE from $0.411px$ to $0.378px$ can be achieved for the MDP algorithm. However, the flow fields of the MDP algorithm are once again erroneous, due to the saturated highlights in the 8-bit images (see above for a detailed explanation). Therefore, I refrain from assessing the results further. The BTF reflectance model reduces the mean DFE from $0.364px$ to $0.249px$ for the PPB algorithm and from $0.451px$ to $0.239px$ for the H&S algorithm. In contrast to the *Top View*, the BTF3K model improves upon the BTF2K model with an additional reduction of the mean DFE from $0.249px$ to $0.204px$ for the PPB algorithm and from $0.239px$ to $0.223px$ for the H&S algorithm. For the TEX model, a similar improvement cannot be observed.

The large standard deviations in all combinations of reflectance models, texture resolution and flow methods indicate that significant outliers exist in all cases. Therefore, no model seems to simulate this sequence with close to perfect accuracy.

The outliers can be better assessed by the DFE distributions depicted in Figure 4.11. For the TEX2K model the distribution is very similar to the *Top View* distributions. It is however slightly broader with few more errors below $0.05px$ and a few more errors above $1.5px$. In contrast to the *Top View*, the increased texture resolution of the TEX3K reduces the errors slightly. There are notable less DFE in the $[0.2, 0.5]$ range, but the large outliers are essentially the same. However, more than 45% of the pixels have DFE above $0.2px$ and most certainly include some errors induced from the reflectance model.

The DFE can be significantly decreased using the BTF re-

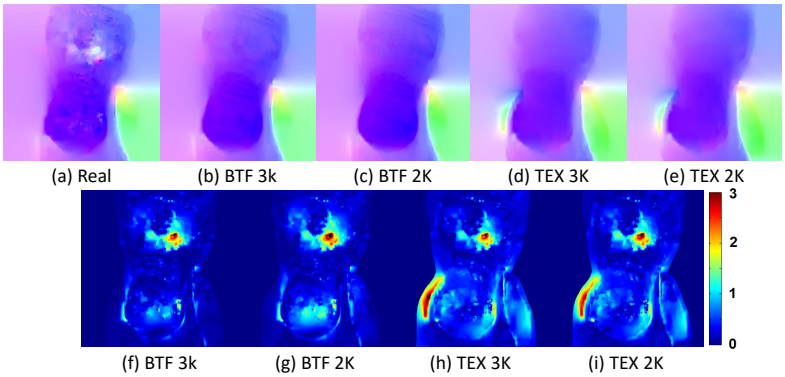

Figure 4.12.: Flow fields of the specular frontal sequence computed with the Papenberg Method. Strong specular effects are present at grazing angles of the object, *i.e.*, the object boundaries. The optical flow algorithm cannot resolve enough detail and the flow fields of the real-world images (c) are blurred across the object boundaries. This can be replicated with the BTF models (a) and (b) but not with the TEX models (d) and (e). Discontinuities of the real-world flow fields (c) occur around the forehead, where mirror-like reflections are combined with delicate geometric detail. The discontinuities cannot be replicated with either method. The corresponding difference flow errors are visualized in (f)-(i). The errors in (h) and (i) at the object boundaries are reduced to almost zero with the BTF models in (f) and (g). They are decreased at the forehead but remain abundant. This matches the observation of the perceptual analysis, that the angular resolution limits the BTF model accuracy in extreme cases of mirror-like reflectivity with delicate geometric details. The higher-resolution variants can reduce the error overall, but not the most critical errors around the forehead.

flectance model. However, in contrast to the *Top View* sequence some more substantial errors remain. For the BTF2K model, roughly 14% of the pixels have DFE values higher than $0.5px$, indicating that the reflectance model causes flow field errors. Out of those more than 5% are larger than $1px$ most certainly suggesting significant errors from the reflectance models. With the BTF3K model, the pixels with DFE values higher than $0.5px$ amount to roughly 11% of the total pixels and the pixels with errors larger than $1px$ are reduced to 3%. This is a significant reduction of outliers from the higher-resolution BTF.

Most pixels, *i.e.*, almost 70% for BTF2K and more than 75% for BTF3K, have low subpixel errors of less than $0.2px$ that fall within the accuracy of the flow algorithm. Only a few pixels fall in the $[0.2, 0.5]$ range where only minor errors are possible invoked by the reflectance model. This indicates that most pixels are simulated with close to perfect accuracy and that some large errors are responsible for flow differences in only a few regions.

Figure 4.12 visualizes the differences in the computed flow fields. The donkey boundary (bottom left of the images) is blurred, due to specular grazing effects and inter-reflections. This translates to errors in the real-world flow fields. The edges are not distinguishable enough for the PPB algorithm Thus the flow field is blurred across the object boundary. The TEX2K and TEX3K model produce more pronounced object boundaries and cannot model the strong specular effects at grazing angles. As a result, the optical flow algorithm can resolve the object boundary. The DFE for the TEX2K and TEX3K model is high in this area, forming a large region with DFE values of $2 - 3px$. BTF2K and BTF3K models can replicate this effect and do not exhibit large DFE values in this area.

The donkey head is highly reflective forming medium specular highlights at the nose (bottom of the image) and a strong spatially varying highlights at the delicate structures of the donkey forehead (top of the image). An additional medium specular highlight is visible at the donkey torso (bottom right of the image) The strong highlight at the forehead causes discontinuities in the real-world flow fields.

The TEX model fails at recreating this critical effect and therefore produces flow fields that are smooth in all areas of the depicted donkey statue. Consequently, the DFE values are plentiful in this area, forming a region with errors of $2 - 3px$.

Both medium specular highlights can be replicated with the BTF2K, and BTF3K models and the DFE are significantly reduced in the corresponding areas. The highlights at the forehead are fainter, and the spatial variations are smoothed out by the BTF models. Consequently, the DFE values are large in this area, forming a region with errors of $2 - 3px$ similar to the TEX models. This matches the previous hypothesizes, that the BTF reflectance model is generally capable of simulating the critical reflectance effects, but is limited by its angular and geometric sampling when mirror-like reflections are combined with delicate geometric details. The increased spatial sampling of the BTF3K model can reduce the overall differences in the flow fields, but does also fail to simulate this most challenging case to the full extent.

4.5. Summary and Outlook

The work presented in this chapter analyzes the reflectance model quality in the context of optical flow performance analysis. To this end, two high-quality reflectance fields were captured using a sophisticated dome setup: One for a mostly diffuse clay pyramid with small geometric details and some subsurface scattering and one for a brass donkey statue with varying degree of specular reflectivity and delicate geometric details. Four real-world sequences were captured and synthetically replicated with two reflectance models: i) The sampled reflectance field, which is stored in a bidirectional texture function. ii) The simple Lambertian BRDF, which is often used for synthetic optical flow datasets and acts as a reference. Three well-known optical flow algorithms were applied to the real sequences and their synthetic

counterparts.

I assessed the quantative perceptual differences using the structural similarity metric [169]. In contrast to the PSNR and MSE, this metric is better suited for an image quality assessments of optical flow datasets, as it focuses on the image structures which in turn affect optical flow performance. The renderings of the Lambertian BRDF cannot replicate critical reflectance effects for optical flow performance analysis, such as the specular highlights of the donkey statue and the shadowing and subsurface scattering of the pyramid fissures. By contrast, the reflectance field renderings are indistinguishable from the real scenes, to the human eye. Close-up views, however, reveal small differences between the reflectance field renderings and real-world sequences. While in general all critical reflectance effects can be replicated, the reflectance field renderings lack small spatial variations of specular details, where mirror-like reflectivity is combined with delicate geometry. These differences can be attributed to the angular and spatial sampling of the reflectance fields, the resampling and SVD compression of the BTF format and the geometric sampling of the object shape.

Therefore, increased sampling rates and less aggressive BTF compression can further increase the reflectance model quality, to the point where even the most challenging highlights could be simulated to the full extent. In the used dome setup, all sampling rates are fixed except for the spatial resolution of the reflectance field, which yielded slight improvements of the perceptual quality.

I quantified the impact of the reflectance model quality by using a difference flow error metric. It measures how close two flow fields are for presumably identical (real and synthetic) scenes, similar to the establish endpoint error. The findings show that for smooth and diffuse surfaces, accurate reflectance field renderings are not needed and simple material models are good enough. The average difference flow errors between real and synthetic flow fields are low and fall well within the accuracy of the optical flow algorithms for both reflectance model types. When mostly diffuse materials boast small geometric details and subsurface scattering, the reflectance field renderings hold a notable advantage over the Lambertian BRDF model renderings and perform with close to perfect accuracy.

For more challenging metallic surfaces the Lambertian BRDF

model performs poorly. Significant difference flow errors occur even for faint specular highlights and mostly flat geometry. By contrast, reflectance field renderings are clearly better: The difference flow errors can be significantly reduced in areas featuring medium to high specular highlights, at grazing angles and for medium geometric details. However, the observed visual discrepancies for mirror-like reflections at delicate geometric structures translate to differences in the flow fields. The real-world flow fields boast discontinuities under these conditions, which is not replicated to the full extent in the reflectance field flow fields. On the other hand, the impact on the optical flow results is so small, that in all cases more than 95% of all evaluated pixel locations are less than 0.5 pixels different in flow magnitude.

We can thus conclude that reflectance field renderings are generally a valid tool for performance analysis of optical flow and provide a significant improvement over commonly used material models. The visual differences and errors in the flow fields start to increase for highly specular objects with lots of fine geometric detail on the mesoscopic pixel scale. In those cases, a denser angular sampling of the reflectance fields and finer reconstruction of the object's shape, should be considered.

5

Acquisition and Benchmarking of BRDF Models For Synthetic Optical Flow Datasets

This chapter is based on my work previously published in [46]

5.1. Motivation

Reflectance fields are an accurate representation of the real-world reflectance of surfaces and objects. They can be sampled with sophisticated measurement devices resulting in a large amount of reflectance data, *i.e.*, 50GB for a small $4 \times 4cm$ patch. Typically, they are then resampled and compressed with linear basis decomposition methods into a bidirectional texture function (BTF). The study in chapter 4 showed that measured BTFs rendered with unbiased path tracing facilitate accurate synthetic datasets, which in most cases, optical flow algorithms cannot distinguish from real-world datasets.

However, there are some limitations: BTFs sample the actual reflectance of surfaces in discrete steps. The reflectance values for intermediate angles are then linearly interpolated. As depicted in Figure 5.1, this can lead to an inaccurate and fainter modeling of very sharp mirror-like reflections and increased Fresnel reflectivity towards grazing angles. By contrast, BRDFs can overcome this limitation by performing model-based interpolation on the measured samples and can more accurately represent the real-world reflectance. Besides, BTFs have other limitations concerning rendering and virtual scene creation.

Modern rendering systems are typically built on analytical BRDF models with a clear separation of diffuse and specular reflectance, *i.e.*, a separation that conforms to how light interactions occur in the real world. This separation is useful for importance sampling and the efficiency of the simulation. On the contrary, BTFs do not follow that separation and are, there-

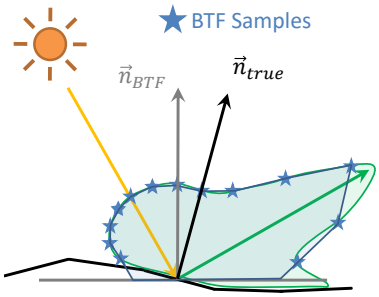

Figure 5.1.: BTFs (blue) discretely sample the real-world reflectance (green) of surfaces. They are in general accurate enough for optical flow performance analysis with some limitations. As they perform linear interpolation of samples, they can miss strong specular reflectance peaks and Fresnel reflections. In addition to being better supported and more efficient, BRDF models can better simulate these effects, due to model-based interpolation. Thus, they are promising candidates for optical flow performance analysis.

fore, often not supported. For the few systems that can handle BTFs, the compressed BTF has to be decomposed for each light surface interaction of the simulation. Compared to analytical BRDF models, the shading costs are substantially increased, and the BTF cannot be importance sampled with the same efficiency. For large complex scenes, which consist of many different materials, the BTF data size is also a limiting factor. Consequently, the quality and diversity of reflectance textures have to be reduced to fit on the GPU memory.

While capturing the surface reflectance of scene objects can lead to high confidence datasets, it is in reality hardly feasible for a large number of complex scenes. Example based texture synthesis and interpolation methods are promising methods to create high confidence datasets from a few measured materials. While both synthesis and interpolation are possible [128], the higher dimensionality of the BTF data leads to additional difficulties and limitations compared to analytical BRDF models.

The work highlighted in this section aims to answer the following questions: One, can we derive more convenient analytical BRDF models from reflectance field data, which are as good as the BTF representation? Second, how accurately do we have to model surfaces and surface reflectance for synthetic optical flow datasets, depending on surface type? In contrast to the work in chapter 4, the study is based on synthetic images alone. A purely synthetic approach has the advantage of facilitating accurate ground truth and the ability to create randomized scenes with a diverse set of reflectance effects, which can hardly be achieved in a lab. This approach is valid, as the BTF reference data is proven to be adequate in the context of optical flow algorithms when using materials without mirror-like reflections.

The contents of this chapter are based on the findings of the previous Chapter 4. The work presented here is solely my contribution unless stated otherwise and included for the sake of readability.

5.2. Contributions

The contributions of the work presented in this chapter are threefold: First, I propose a method that can create highly accurate, spatially consistent BRDF model fits from BTF data, preserving even small geometric details on the mesoscopic texel scale.

Second, I present new insights on five reflectance effects in the context of optical flow performance analysis: *specular highlights*, *anisotropic reflectance*, *large scale shadow-masking with subsurface scattering*, *fine geometric structures* and *Fresnel reflectivity*. Third, I set the model requirements for discriminating synthetic optical flow datasets: I show that for all five effects an efficient BRDF model can substitute the BTF, creating synthetic images with virtually identical optical flow behavior. The best performance can be achieved with a physically based anisotropic Cook-Torrance model and a novel normal estimation.

Fitting analytical BRDFs to BTF data is complex: First, the reflectance is only sparsely sampled in the angular domain. Second, only a coarse geometry is provided, and the remaining geometric details are encoded within the reflectance information. When fitting BRDF models without physical constraints, the textures are spatially inconsistent and fine structures are blurred, due to parallax artifacts and violations of the normal incidence assumption of BRDF models. To facilitate spatially consistent BRDF fits, a constraint fitting process is proposed. This process can be extended with a normal estimation preprocess to preserve the mesoscopic geometric details of the BTF.

I show that by utilizing the proposed methods, the appearance of the BTF reference can be closely matched in most cases. High-frequency highlights and shadow-masking effects around *fine geometric structures* are preserved. Minor differences occur for sharp *specular highlights*, where the BRDF model reflections more intense, due to a superior model based interpolation of the measured samples. More notable differences remain for surfaces *large scale shadowing-masking and subsurface scattering effects*, which cannot be simulated with either BRDF model.

The conducted optical flow experiments reveal insights on the required BRDF model accuracy for the five reflectance effects:

i) *Large scale shadowing-masking and subsurface scattering effects* do not present a challenge for optical flow algorithms, as they are mostly diffused and motion-invariant. As a result, differences between the computed flow fields are minimal, and all BRDF model fits can be used with high accuracy.

ii) As expected, *specular highlights* affect optical flow algorithms the most, due to the highly motion-variant nature. All BRDF models with normal estimation applied produce flow field errors that are virtually identical to the BTF model flow field

errors.

iii) Surfaces with *anisotropic reflectance*, *e.g.*, fabrics and cloth, are not generally more challenging for optical flow algorithms than isotropic surfaces. However, anisotropic BRDF model fits are required, since isotropic models tend to encode the anisotropic reflectance in the diffuse term, resulting in significantly smaller flow errors.

iv) Surfaces with *fine geometric structures* are challenging to optical flow algorithms, due to high-frequency shadow-masking and highlights. BRDF models with normal estimation are needed to match the BTF reference flow fields. When combined with *anisotropic reflectance* or increased *Fresnel reflectivity*, only a physically based Cook-Torrance model with normal estimation performs with almost perfect accuracy.

v) Coated surfaces with intense *Fresnel reflectivity* are only challenging when combined with other effects, *e.g.*, the fine geometric structures of leather. In that case, additional normal estimation is required to match the BTF reference flow fields. For all coated materials, a physically based Cook-Torrance model performs notably better than the empirical Ward and Phong models.

5.3. Creating BRDF Reflectance Data For Optical Flow

Sophisticated dome setups can be used to accurately measure the reflectance fields of surface and objects [132]. The recorded reflectance data often exceeds several hundred GB and is therefore not suited for storage and rendering. Typically, linear basis decomposition is applied to the reflectance field data to reduce the data size of the BTF and speed up the rendering process (*c.f.*, section 2.4.3. They are data efficient and can achieve almost lossless quality.

In this section, I first present the chosen reference data, consisting of material samples with a large variety of reflectance effects. Then I show how the BRDF model fits are formulated as a non-linear least square fit of the ABRDF representation of the BTF data. I present the encountered problems of an unconstraint fit, due to the sampling properties of the underlying BTF data. Finally, I propose a constraint fit with normal estimation, that can better deal with the BTF sampling.

5.3.1. Reflectance Field Reference Data

To study the required reflectance model accuracy, in the context of synthetic optical flow datasets, a reference dataset with sufficiently accurate reflectance field data is needed. The dataset should include a wide range of reasonably large material samples and sufficiently accurate sampling of the geometry and the reflectance in the spatial and angular domains. Ideally, it would facilitate uncompressed reflectance fields for optimal BRDF fitting as well as compressed BTF data. Unfortunately, no such dataset exist and the compressed BTF data has to be used for BRDF model fits. Consequently, the BRDF models always include artifacts induced from the BTF format and the format with the least aggressive compression creates the best results.

The BTF database Bonn meets the requirements except for the availability of an uncompressed reflectance field format [156]. It contains reflectance field measurements for 84 planar material samples for seven semantic classes, *i.e.*, 12 samples each for carpets, fabrics, felts, leathers, stones, wallpapers, and woods. The reflectance field data of the material samples are captured with the Dome1 setup of the University of Bonn. For a detailed description of the measurement process, I refer to the work by Schwartz *et al.* [132]. It is encoded in a bidirectional texture function (BTF) for 151×151 distinct view and light directions, which are evenly distributed on the hemisphere above the samples. The BTF data comes in two formats: i) A Decorrelated Full-Matrix-Factorization (DFMF), directly sampling the reflectance field measurements. ii) A Full-Matrix-Factorization (FMF) resampled with respect to the measured height map. The FMF format has the advantage of including accurate silhouettes of the surface geometry and increased detail for structured surfaces that deviate from normal incidence. The quality of the FMF format closely resembles the quality of the BTF format used in Chapter 4. It has a slightly more aggressive compression in the spatial domain and a more accurate compression in the angular domain. As the angular sampling is the most limiting factor in terms of perceptual and optical flow simulation quality, it is a suitable format for this study. The chosen BTF dataset has proven applicability in the computer vision domain, as it was captured for and successfully applied to synthesize training images for a material classification network [156].

From the available materials samples I chose a subset of seven,

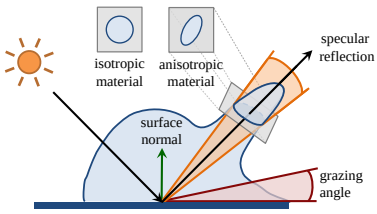

Figure 5.2.: Real-world reflectance is partly diffused in all directions and partly distributed around the specular reflection direction and the grazing angles, *i.e.*, low viewing angles. The specular distribution can be isotropic or anisotropic, *i.e.*, the reflectivity depends on the azimuth angle of the reflection direction. All BRDF models consist of a view independent Lambertian diffuse term and a view-dependent model specific specular term. The empirical Phong model can only model isotropic reflections without Fresnel reflectivity at grazing angles. The empirical Ward model can additionally model anisotropic reflections. The physically based Cook-Torrance can model anisotropic reflections and Fresnel reflectivity at grazing angles.

covering each material class of the dataset and a large variety of reflectance properties: *felt05*, *wood07*, *fabric07*, *leather01*, *stone06*, *wall02* and *carpet02*. For clarity, I omit using the numbered suffix in the following sections. Since all seven classes are dielectrics, I additionally used an *aluminum* and a *copper* BTF, which was measured with the same dome setup, to include metallic reflectance properties. This set of materials covers a broad spectrum of material reflectance properties: almost perfectly diffuse reflectance with subsurface scattering (*felt, carpet*), metallic specular reflectance (*aluminum, copper*), anisotropic specular reflectance (*fabric, leather, wood*), smooth coatings and strong Fresnel reflectivity (*stone, leather, wood*) and surfaces with pronounced geometric details (*leather, carpet, felt, wall*). Using these samples, I can create various critical reflectance effects for optical flow algorithms and assess the required reflectance model quality for synthetic optical flow datasets.

5.3.2. Fitting BRDFs to BTF Data

BRDF models are efficient analytical representations of reflectance properties of surfaces and objects 2.4.2. As such, they are tied to BTF compression methods, which are concerned with reducing the vast amounts of data from reflectance field measurements.

The Lafortune lobe model [75] is an empirical, analytical BRDF model, which consists of a finite sum of arbitrarily aligned cosine lobes and was specially designed for fitting reflectance data. The model was applied to BTF compression by McAllister *et al.* [96]. However, the compression is limited in quality for non-homogeneous surfaces compared to linear basis decomposition methods, which are typically used for BTF compression. Additionally, compared to other analytical BRDF models, the model is computationally inefficient, lacks physical plausibility, and most importantly, lacks support in modern rendering and texturing systems. It is therefore hardly used for BTF compression and not further addressed in this study.

On the contrary, other analytical BRDF models are better suited for rendering tasks and de facto standard opaque material models in computer graphics. In this study, three commonly used BRDF models were fitted to the BTF data and analyzed with respect to optical flow performance analysis: a physically plausible modified *Phong* model [74], a *Ward* model with the Duer normalization fix for specular lobes and grazing angles [39]

and a modified *Cook-Torrance* model [20] with an anisotropic microfacet distribution from Ashikhmin and Shirley [5]. The different properties of the BRDF models are outlined in Figure 5.2. All models consist of a lambertian diffuse term and a model specific specular lobe:

$$M(\omega_i, \omega_o, d, s, \alpha) = d/\pi + K(s, \omega_i, \omega_o, \alpha) \tag{5.1}$$

where ω_i, ω_o are the incident and outgoing light directions, d is the diffuse color, π is the Lambertian normalization factor, K is the specular reflectance term, s is the specular color, and n are the model specific roughness or glossiness parameters. For the *Phong* model α corresponds to the one-dimensional Phong exponent of the isotropic cosine lobe, for the *Ward* model α corresponds to the two roughness parameters along the tangent and bitangent directions, and for the *Cook-Torrance* model α corresponds to the two specular exponents along the tangent and bitangent directions. For the exact analytical form of each specular BRDF term, I refer to Section 2.4.2 and the original papers.

To create spatially varying BRDF textures from BTF data, in the least squares sense, the BTF data has to be converted into a format that is compatible with the BRDF. This format is the apparent BRDF (ABRDF) which encodes the reflectance of a BTF texel for a fixed amount of distinct view and light directions, *c.f.*, Section 2.4.3. Typically, those directions are evenly distributed on the hemisphere above the texel. In the used BTF data, the ABRDFs are of size 151×151 and correspond to an average angular resolution of $9.4°$ in both view and light directions [132]. In the normalized format, *i.e.*, the ABRDF is normalized to the irradiance incident on the captured surface geometry (*c.f.*, equation 2.16), the ABRDF is essentially a sampling of the BRDF. However, it additionally models mesoscopic geometry variations and does not follow Helmholtz Reciprocity (*c.f.*, section 2.4.1), *e.g.*, a texel can include subsurface scattering light paths from other parts of the surface.

For each BTF Texel I constructed the minimzation problem as a least square fit of the BRDF model paramaters $p = (d, s, \alpha)$:

$$p = \underset{p}{\operatorname{argmin}} \frac{1}{N} \sum_{i,o} (\cos\theta_i \cos\theta_o (\hat{A}_x(\omega_i, \omega_o) - M_x(\omega_i, \omega_o, p)))^2 \tag{5.2}$$

where x is the BTF texel, $\hat{A}$ is the normalized ABRDF of the texel consisting of the measurements for ω_i, ω_o the incident and outgoing light directions, θ_i, θ_o are the corresponding elevation angles and $M_x(\omega_i, \omega_o, p)$ is the reflectance value of the texels BRDF model for the incident and outgoing directions as in equation 5.1.

With a decreased camera elevation, a smaller portion of the surface texel is visible from the camera. This portion is proportional to the cosine of the elevation angle $\cos\theta_o$. Respectively, the incident irradiance of the light source on the surface texel is proportional to the cosine of its elevation angle $\cos\theta_i$. Therefore, the uncertainty of the ABRDF measurements is inverse proportional to the elevation of both the camera and the light, and the $1/(\cos\theta_i \cos\theta_o)$ factor of the objective function is mathematically motivated.

5.3.3. Encountered Challenges of Unconstraint Fits

Fitting analytical BRDF models to measured reflectance data has been successfully applied before, *e.g.*, in the work of Ngan *et al.* [110] and Matusik *et al.* [93]. They densely captured the reflectance data of spherical homogeneous material samples from a camera and a moveable light source. For a detailed description of the measurement setup, I refer to the previous section 2.4.4.

The captured data consists of more than $4M$ samples, which results in an angular resolution below 1° around the perfect mirror direction. Due to this high angular sampling density and the perfect underlying surface normals, it can be used for direct visualization or BRDF models can be fitted to the data with relative ease, using non-linear squares optimization methods. However, the data is limited to the reflectance of a single surface point, by design. Therefore, it can only be used to model homogeneous flat material reflectance, *e.g.*, smooth metals and paints. The BTF bypasses this limitation by sampling the entire reflectance field and storing it in a texture. It includes spatially varying reflectance and mesoscopic detail, *i.e.*, geometric details on the texel scale, and is suited to model the reflectance properties of more complex heterogeneous materials, which are relevant for synthetic optical flow datasets. However, the specific sampling properties of the BTF model increase the difficulty of the BRDF fits in comparison to the methods described by Ngan *et al.* [110] and Matusik *et al.* [93]. In the following I describe the encoun-

tered challenges of this process:

Sparse Angular Sampling The MERL BRDF data includes roughly $4M$ reflectance samples, an equivalent to a sub 1° angular resolution. By contrast, the angular sampling density of the BTF data is significantly lower with 22801 reflectance samples corresponding to an angular resolution of 9.4°. The BTF rendering performs a linear interpolation between the reflectance samples. Therefore, mirror-like reflectivity might be insufficiently sampled, *e.g.*, for smooth metals, and some specular details might be missed (*c.f.*, chapter 4. An analytical BRDF fit can potentially better model mirror-like reflectivity, as a BRDF model interpolation can be considered superior to a simple linear interpolation. However, the coarse angular sampling can result in spatially inconsistent specular terms using unconstraint nonlinear least squares optimization, *i.e.*, some texels might include severe specular reflectivity, and other potentially neighboring texels might have fainter specular reflectivity. For almost diffuse materials, a similar problem can occur. The coarse angular sampling can create ambiguities of BRDF terms, *i.e.*, there might not be a precise distinction between a rough specular reflectance and a diffuse Lambertian reflectance. As a result, one can observe spatial inconsistency for very diffuse materials, where some texels might have a stronger rough specular term and a weaker diffuse Lambertian term or the other way around. The spatial inconsistencies in both cases can be amplified when high fidelity mesoscopic detail is present, a BRDF fitting challenge that I explain subsequently.

Mesoscopic Detail The reflectance data of the MERL BRDF database is sampled for a single surface point with a perfect surface normal, *i.e.*, a surface that points upwards along the normal incidence. Hence, the data is well suited for BRDF model fits, that assume a normal along normal incidence. By contrast, the BTF material samples are $4cm \times 4cm$ surface areas with detailed geometry and surface normals that deviate from normal incidence. A structured light scanner coarsely scans the surface geometry, and the BTF data is resampled with respect to that geometry. However, the coarse geometry does often not match the real-world geometry on the mesoscopic texel scale, *c.f.*, Section 2.2.2. These mesoscopic geometry details are encoded within the BTF reflectance data. Later results show that

for many material samples, the real-world normals often deviate from the measured geometry normals by up to 30°. BRDF models do, however, assume normal incidence, causing inaccuracies for unconstraint non-linear least squares fits of the BTF data. The mesoscopic details are blurred out in the fitted textures and structures are lost. Parallaxing artifacts are caused at low viewing angles, where the real-world reflection direction points below the coarse geometry. The spatial inconsistencies caused by sparse angular sampling are further amplified for material samples with lots of mesoscopic details, such as the *leather*, *fabric* and *carpet* samples.

Impact on Optical Flow Unconstraint least square fits of the BRDF models can result in a spatially inconsistent appearance, that is detrimental to optical flow performance analysis. New structures are induced from the BRDF models, affecting the pixel correspondences of the optical flow algorithms. The loss of mesoscopic details in the BRDF fits could affect the performance of optical flow algorithms, which often rely on such mesoscopic structures.

Therefore, additional optimization steps and constraints are needed to ensure that the BRDF model fits are creating a more consistent and detailed visual appearance and are thus better suited for optical flow performance analysis.

5.3.4. Improved BRDF Fit to BTF Data

In this section, I detail my proposed method to create visually pleasing and consistent results when fitting efficient BRDF models to BTF reflectance data. This method accommodates for the challenges due to the sparse angular and geometric sampling of the reflectance fields and captures the encoded mesoscopic geometric details of the BTF data. It is applicable to arbitrary, mostly opaque material samples with varying reflectance properties, *e.g.*, the chosen material samples from the BTF database Bonn for the following experiments.

General Setup In all BRDF model fits two specular lobes were used. While a second specular lobe is not physically correct in all cases, *i.e.*, only coated materials have two physical interfaces and thus two specular terms, it can improve the quality of the fits for two reasons: i) Real-world materials often have broad specular

distributions, *i.e.*, long tail off-specular reflections. Most specular distributions do not sufficiently model off specular reflections and a broad second specular lobe was suggested by Ngan *et al.* [110] and Cook and Torrance [20] to improve the quality of the fits. ii) For structured material with high-frequency mesoscopic detail encoded in the BTF, a second specular lobe can help to capture this detail better. A $\log(1+\frac{s}{4})$ Cauchy loss function was used for the least-square optimization of the BRDF models. The negative impact of outliers, present near grazing angles and due to the mesoscopic geometric details in the BTF, is therefore reduced.

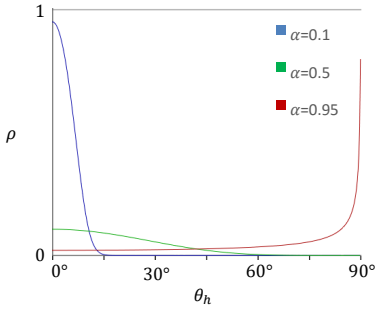

Figure 5.3.: The proposed fitting approach utilizes two independent specular lobes. The first specular lobe with roughness value between 0.1(blue) and 0.5(green) is used to model the main specular reflections, which is mostly distributed for half angles θ_h close to the surface normal. The second specular lobe uses roughness values between 0.5(green) and 0.95(red) and mostly models the off-specular reflections and increased Fresnel reflectivity for flat half angles. The lobes are fitted in a two-step process, which ensures that the diffuse term is spatially consistent.

Consistent Fitting Process

A two-step fitting process was used applying knowledge about physical reflectance properties and limitations of the underlying BTF data. This process helps to stabilize the BRDF fits in the spatial domain and make them visually consistent.

Initial Diffuse Estimate In the first step, an initial estimate of the diffuse reflectance parameters d_0 is computed from the view and light directions around normal incidence, using a single narrow specular term for the main specular reflection component. More precisely, equation 5.2 is solved with $p = (d_0, s_0, \alpha_0)$ for elevation angles $\theta_i, \theta_o < 45°$ and a single specular term with constraint parameter bounds for the roughness parameter $\alpha_0 \in [0.1, 0.5]$. The maximum elevation angles of the view and light directions are chosen to ensure that the increased Fresnel reflectivity at low grazing angles does not influence the diffuse reflectance estimate. The roughness parameter bounds were chosen for the *Ward* and *Cook-Torrance* models and are equivalent to a *Phong* glossiness exponent parameter range of $[6, 202]$. The corresponding specular distributions for the *Cook-Torrance* model are depicted in Figure 5.3. It visualizes the specular reflectance for the minimum and maximum roughness value with respect to the angle between the half vector and the surface normal. The minimum roughness of 0.1 is chosen to ensure that the sharpest specular highlights can still be sufficiently sampled with the 9.4° angular resolution of the BTF Data. The maximum roughness value of 0.5 ensures that the mostly diffuse reflectance properties are modeled with the Lambertian diffuse term of the model and not the main specular lobe.

Full Estimate In the second step a larger portion of the view and light directions is utilized, and the specular lobe of the previous step is replaced by two independent specular lobes, one for the main specular reflections and one for off-specular reflections. At the same time, the diffuse estimate d_0 from the previous step serves as parameter bound for the finale diffuse reflectance estimate d. More precisely, equation 5.2 is solved with $p = (d, s_{main}, s_{off}, \alpha_{main}, \alpha_{off})$ for elevation angles $\theta_i, \theta_o < 60°$ with the diffuse reflectance parameter $d \in [k \cdot d_0, d_0]$, the main specular lobe with $\alpha_{main} \in [0.1, 0.5]$ and the off specular lobe with $\alpha_{main} \in [0.5, 0.95]$. By utilizing more view and light directions, the increased Fresnel reflectivity can be better modeled. However, the ABRDF are increasingly inaccurate for low angles, and thus it was beneficial not to use the full direction set with elevation angles up to 75°. The upper bound for the diffuse reflectance corresponds to the initial estimate. It ensures that the increased Fresnel reflectivity near grazing angles does not affect the final diffuse estimate, and is instead correctly modeled by the specular terms. The downscaled lower bound is used to clear space for the off-specular lobe. I empirically chose $k = 2/3$ for visually pleasing results. The parameter bounds for the main specular lobe were chosen as in the previous step. The parameter bounds for the off-specular lobe were chosen to cover specular reflectance effects in between the main specular lobe and the diffuse reflectance term.

Normal Estimation

The surface geometry of the material samples is only coarsely sampled in the BTF data. The remaining geometric details on the mesoscopic texel scale are encoded within the BTF reflectance. The previously proposed BRDF fitting process is spatially consistent, but cannot model these mesoscopic details. Therefore, I propose a normal estimation preprocessing step that can accurately model the present geometric details. It can be assumed that the highest reflectance occurs along the perfect mirror direction, which is directly dependent on the surface normal. Extending ideas from Shape from Shading [122] and Photometric Stereo [41] the normal can be estimated. However, instead of assuming diffuse reflectance, the density of the BTF data allows the incorporation of a specular term for more accurate surface normal estimation. The optimization function from equation 5.2

is solved for a normal estimation model with an anisotropic cosine lobe around the perfect mirror direction:

$$M_x(\omega_i, \omega_o, \phi, n, d, s, n_u, n_v) = \frac{\omega_i \cdot n}{cos\theta_i}\left(\frac{d}{\pi} + s\frac{\sqrt{(n_u+2)(n_v+2)}}{2\pi}(n \cdot h)^{n_u \cos^2\phi + n_v \sin^2\phi}\right) \quad (5.3)$$

where $h = ||\omega_i + \omega_o||$ is the half vector between the view direction ω_o and light direction ω_i, n is the surface normal and n_u and n_v are the cosine exponents along tangent ϕ and its corresponding bitangent direction. The factor $\frac{1}{\cos\theta_i}$ corresponds to a ABRDF without normalization to incident irradiance on the scanned coarse geometry. Instead, the BRDF model is attenuated with the incident irradiance of the estimated normal, modeled with the factor $\omega_i \cdot n$. The diffuse term d/π is a regular Lambertian diffuse term and conforms to the terms of Shape from Shading [122] and Photometric Stereo [41]. The specular lobe is equivalent to an anisotropic Phong Model with a parametrized normal. A standard isotropic Phong model was tested but deviated for anisotropic material samples. The objective function was minimized for $\theta_i, \theta_o <= 60°$. The resulting surface normal correction term can then be used as the input normal for previously presented two-step BRDF estimation. The resulting BRDF models are then both spatially consistent and include mesoscopic geometric details.

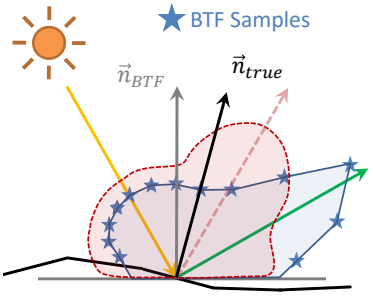

Figure 5.4.: Schematic figure of mesoscopic detail loss with BRDF model fits. The BTF (blue) encodes the surface geometry and surfaces normal (grey) in a coarse mesh. The true surface and its normal (black) can often deviate for surfaces with fine structures. This mesoscopic geometric detail is typically encoded in the BTF samples. BRDF model fits(ref) to the BTF data, which utilize the BTF normals, are thus more diffused and exhibit broader specular highlights. This results in a loss of structure and fine geometric details. An estimation of the true normal can help overcome this issue.

5.4. Experimental Setup

The methods described in the previous section 5.3 were used to fit three analytical BRDF models to the BTF data: a physically plausible modified Phong model [74], a Ward-Duer model with a normalization fix for specular lobes and grazing angles [39] and a modified Cook-Torrance model [20] with an anisotropic microfacet distribution from Ashikhmin and Shirley [5]. Two variants each were created: First, a variant created with the proposed consistent fitting process. Second, a variant with additional normal correction applied. The methods were implemented with the open-source Ceres Solver [1] and applied to each material sample. This resulted in a total amount of 54 BRDF

[1] ceres-solver.org

model fits, which are distributed to the nine material samples, three BRDF models, and two BRDF model variants.

To assess the accuracy of the BRDF model fits in the context of optical flow performance analysis, randomized test scenes were created. The sampling of the geometry, camera parameters, and light parameters were chosen to create a large variety of camera and light motions as well as motion-variant reflectance effects that challenge optical flow algorithms, *e.g.*, specular highlights with varying size and motion. The BTF and BRDF materials were applied to each test scene and rendered using unbiased path tracing. The resulting synthetic images were then used as input for two optical flow algorithms: The *MDP-Flow2 Algorithm* by Xu et al. [163], it achieves high performance levels on the Middlebury dataset and comes with a published implementation, and a multi-scale, non-linear version of the classical *Horn&Schunck algorithm* [62], which was implemented with the open source Charon framework [43]. Parameters for both methods were chosen to produce flow fields as accurate as possible for arbitrary scenes. Finally, the flow fields were analyzed concerning Endpoint Error (EPE) and *Reflectance Aware Endpoint Error* distributions. The latter analysis takes into account the EPE distributions with respect to normals, camera positions, and light positions, enabling the assessment of motion-variant reflectance effects, *i.e.*, specular highlights and grazing effects.

5.4.1. Generation of Randomized Test Sequences

The scanned mostly planar surface geometries of the BTF reference data are encoded in height maps. By using the height maps directly, all synthetic renderings exhibit reflectance effects with similar spatial distributions, *e.g.*, structures and specular highlights are of similar size in each rendering of a material sample. This is unrealistic and unrepresentative for real-world conditions, *i.e.*, real-world surfaces often belong to more complex non-planar shapes. To better asses a large variety of reflectance effects, the height maps were displaced by $1/f^4$ gradient Perlin noise [54], creating naturalistic deformations of the surface geometry. The noise focuses on low-frequency displacements with the size of the material textures, *i.e.*, noise with the size of $4 \times 4cm$, and decreases by a factor of 4 with each subsequent octave. Similar noise functions are often used to procedurally create naturalistic geometry for materials such as cloth, terrain, tarmac and

stone materials.

A pinhole camera was used with a field of view of 50° and a quadratic sensor size of $512 \times 512px$. The camera position was uniformly sampled around the center position of the material surface, with an elevation angle $\theta \in [0, \pi/4]$, an azimuth angle $\phi \in [0, 2\pi]$ and a distance towards the center of the material sample in the range $[0.4, 0.5]m$. The target-vector of the pinhole camera was chosen to point at a random point of the surface geometry, and the up-vector was uniformly sampled in perpendicular directions. The camera was moved on a linear path in 3D space, with a camera distance between image pairs, that was uniformly sampled in the range of $[0.0, 0.01]m$. The camera rotation between the image pairs was sampled with a normal distribution and $\sigma = 0.1°$, favoring small rotations.

As a light source, a sphere with a diameter of $5mm$ and a white light emission with a radiance of $200W/(sr \cdot m^2 \cdot Hz)$ was used, imitating a white LED bulb. A small amount of ambient light was added, *i.e.*, an ambient radiance of $0.3W/(sr \cdot m^2 \cdot Hz)$, to ensure that some light was always visible by the surface. The light extrinsics were sampled in the same fashion as the camera extrinsics. However, the distance to the material sample and the traveled distance between image pairs was increased by 50% to ensure that the light source was not visible by the camera.

A total of 39 test scenes were sampled and then rendered with BTF reference data as well as the BRDF fits of each material sample. To this end, I used the open source, physics-based renderer "Mitsuba"[2] with an unbiased path tracing rendering algorithm and a sample count of 1024 samples per pixel. The BTF rendering was performed with a plugin supplied by the creators of the BTF database Bonn [156]. For the Cook-Torrance fits a new Mitsuba plugin was implemented. For the Phong and Ward fits, I used the original plugin from Mitsuba. The ground truth flow was ray traced with the method of McCane *et al.* [98]. Overall the random sampling of the virtual scene and camera extrinsics resulted in average ground truth flow magnitude of $7.4px$.

5.4.2. Reflectance Aware Endpoint Error

The quality of the BRDF fits, and the visual appearance of the synthesized images can give the first indication of system-

[2] www.mitsuba-renderer.org

atic differences between the reflectance models, which can lead to differences in the computed flow fields. However, the main goal of this study is the assessment of the BRDF models concerning different material types and the simulation of critical reflectance effects for optical flow performance analysis. Optical flow is sensitive to local brightness changes over time, *i.e.*, motion-variant reflectance effects such as moving specular highlights. However, it is robust towards brightness differences that are motion-invariant, *e.g.*, differences in the diffuse reflectance. Consequently, the errors around motion-variant reflectance effects are of primary interest.

The Endpoint Error (EPE) computes the average magnitudes of the differences between the ground truth flow fields and the computed flow fields [114]. It is an established measure to assess the general performance of optical flow algorithms, assuming ground truth flow is provided. However, the EPE cannot distinguish between critical errors and errors in less critical parts of the scene. In the context of stereo disparity estimation, geometry aware stereo metrics have been established to accurately measure the performance of algorithms for critical parts of the scenes, *i.e.*, fine structures, depth discontinuities and planar surfaces [61].

This approach can be adapted. Fine structures, depth discontinuities, and planar surfaces are not as relevant for measuring the BRDF model quality for optical flow datasets. However, the geometric properties of the surface, the position of the light source, and the position of the camera are essential aspects for the generation of motion-variant reflectance effects. Therefore, I use a reflectance-aware analysis of the EPE distributions. More precisely, the flow vectors are partitioned into intervals according to the angular distance between the perfect light reflection direction ω_r and the view direction of the camera ω_o. I then analyze the simulation quality of specular effects by evaluating the EPE statistics with respect to the reflectance-aware intervals. Accordingly, the angle between the local surface normal n and the view vector ω_o help investigate increased Fresnel reflectivity close to grazing angles.

5.5. Results

In this section, I present the results of the conducted experiments. I used a BTF reference dataset (BTF) consisting of nine material samples: *carpet, copper, alu, fabric, felt, leather, stone, wall and wood.* For each material sample I created six analytical BRDF model fits: A **Ward** model, a Cook Torrance(**CT**) model and a **Phong** model as well as the corresponding BRDF model with normal estimation **WardN**, **CTN** and **PhongN**. I randomly sampled 39 test scenes with the presented workflow from section 5.4. The test scenes were rendered for the nine material samples and the seven reflectance models each, resulting in a total of 2457 synthetic test sequences. Finally, the optical flow was computed using two flow methods, a basic Horn&Schunck (*H&S*) and the more sophisticated MDP-Flow2 (*MDP*) algorithm.

The following analysis is concerned with the BRDF model simulation quality of critical optical flow reflectance effects for different material types. I analyze the quality of the BRDF model fits for the remaining cost of the optimization and the perceptual differences of the synthesized images. Then I briefly report general insights regarding material simulation quality from analyzing the EPE distributions of 750 million flow vectors. Subsequently, I present a detailed analysis of the reflectance aware EPE distributions for critical motion-variant effects: The specular reflections on the copper and aluminum samples and the mesoscopic detailed surface geometry on the leather and fabric samples.

5.5.1. BRDF Model Fit Quality

The per texel cost of the objective function (*c.f.*, equation 5.2) for each BRDF model fit averaged out over all material samples is depicted in Table 5.1. As expected, the BRDF models rank in the order of the physical model accuracy, where the physically based anisotropic **CT** model is best, followed by the data-driven anisotropic **Ward** model and the empirical isotropic **Phong** model. Differences between **CT** and **Ward** are however marginal. The proposed normal estimation pre-process can reduce the cost of the BRDF model fits by $\sim 25\%$ on average, with up to $\sim 50\%$ cost reduction for specular and mesoscopic detailed material samples, *i.e.*, aluminum, copper, and leather, as well as

Table 5.1.: Table depicts the average per texel cost of the BRDF model fits over all material samples and the corresponding relative error for each BTF measurement. BRDF rank in order of **CT**, **Ward**, **Phong**. Normal estimation improves the BRDF fits by roughly 25%.

	CTN	CT	WardN	Ward	PhongN	Phong
cost	6.74	10.43	7.39	11.02	14.84	17.50
σ^2_{rel}	2.9%	4.6%	3.3%	4.9%	6.5%	7.7%

a minimum of $\sim 10\%$ cost reduction for almost diffuse material samples, *i.e.*, felt and carpet. The cost values correspond to a relative variance σ^2_{rel} for each measured reflectance value in the range of 2.9% for the most accurate **CTN** model and 7.7% for the least accurate **Phong** model without normal estimation.

5.5.2. Perceptual Comparison

Close-ups of synthesized images for exemplary test scenes of three material samples are depicted in Figure 5.5. The copper sample features high specular reflectivity and low-frequency mesoscopic details. In the depicted image, a pronounced specular highlight is visible. A high-intensity highlight along the main specular reflection direction can be seen in the center and a low-intensity off-specular reflection towards the border of the close-up. All three BRDF models can replicate the main highlight with almost identical appearance and closely match the appearance of the BTF reference. However, the highlight intensity is slightly increased for all BRDF models, likely due to the inaccurate linear interpolation of specular reflections for the BTF reference. For the BRDF models without normal estimation, the highlight is slightly shifted to the top, due to a systematic offset of the scanned surface normal by $\approx 3°$. The normal estimation can correct this offset and better replicate the main specular highlight. The off specular reflections, *i.e.*, reflected light paths further away from the perfect mirror directions, are better simulated with the more accurate **CT** and **Ward** models compared to **Phong** model, which appears to be slightly too dark in said areas.

For the leather material sample with high-frequency mesoscopic detail and medium specular reflectivity, the fine structures of the surface fissures are significantly blurred for all BRDF models without normal estimation compared to the BTF reference. This can be corrected with normal estimation for all BRDF models, which can retain all the fine geometric details of the BTF. The **CTN** appearance closely matches the appearance of the BTF model in all areas. The **WardN** model to a low extent and **PhongN** to a much higher extend models appear darker

around grazing angles, *i.e.*, at the left side of the close-up, where the angle between the view and surface normal is large. This can be explained by the inability to simulate Fresnel reflections adequately. The fissure structures for the **PhongN** model are more pronounced compared to the BTF, **CTN** and **WardN** models. As the underlying surface normals are identical, this can only be caused by the inability to model anisotropic reflectance, which is in turn encoded into the diffuse term.

The felt material sample is a composite layered material, with almost diffuse specular reflectance, high and low-frequency mesoscopic details and strong subsurface-scattering and shadowing-masking effects. Compared to the BTF reference, the appearance of all BRDF models is more blurry and lacks large scale shadowing and scattering effects. Some of the small mesoscopic details can be corrected with the normal estimation, but substantial differences in the low-frequency details remain. This can be explained: The BTF reference can model large scale shadowing effects as well as the light that travels within the material before it is reflected towards the viewer. The BRDF models cannot replicate these effects, as the light entrance and exit position always fall in the same texels. However, due to the diffused nature, these effects are typically mostly motion-invariant and do not pose challenges to most optical flow algorithms.

5.5.3. General Optical Flow Analysis

Flow fields for image pairs that are not bit-accurate identical always differ in some way, due to the algorithms dependency on the statistical distribution of image intensities. However, synthetic renderings can still be considered useful for optical flow performance analysis, when they exhibit closely matching EPE distributions, *i.e.*, the probabilities to create specific error ranges is identical. In the following paragraphs, I briefly report the analysis on the EPE distribution for both flow methods and all nine materials and then detail the most salient material specific EPE distributions. While EPE distributions are not suited to localize the circumstances that lead to systematical differences in the flow fields, they can give us the first indication if and how the BRDF model flow fields differ from the BTF model reference.

MDP EPE Distributions Figure 5.6 depicts boxplots of EPE distributions for the *MDP* flow algorithm for all reflectance mod-

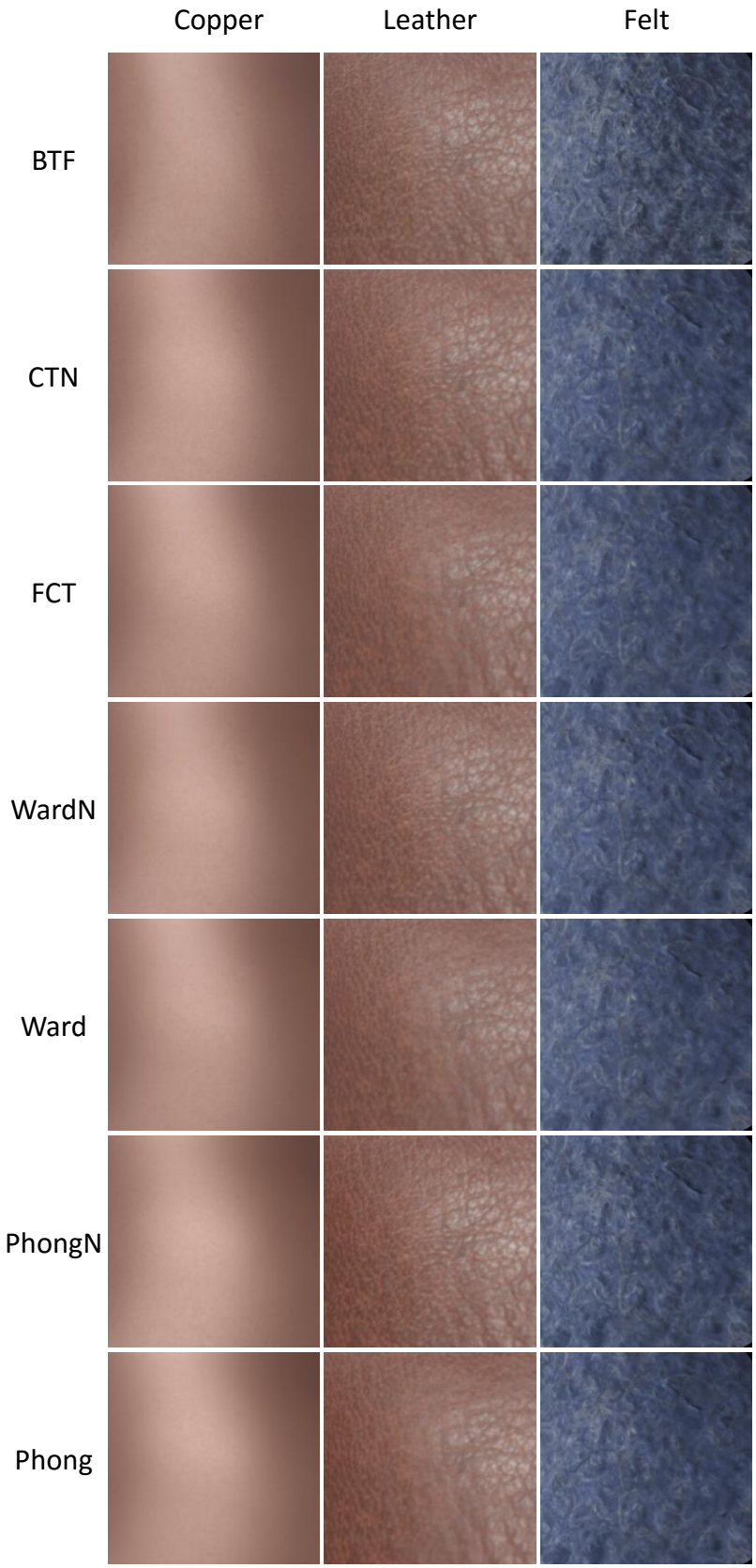

Figure 5.5.: Close-up images of exemplary synthesized images for three material samples and all reflectance models. In the BTF reference image of the **copper** sample (left), a large highlight is visible with high intensity main specular reflections in the center and lower intensity off-specular reflections to the sides. The **CTN**, **WardN** and **PhongN** models can replicate the highlight. However, the main specular reflections are more intense, due to the improved model-based interpolation of the BTF samples. The off-specular reflections are decreased for the **PhongN** model. The **CT**, **Ward** and **Phong** model highlight is slightly shifted to the top, due to a systematic normal offset of $\approx 3^\circ$ in the scanned geometry. In the BTF reference image of the **leather** sample (middle) high-frequency specular highlights are visible to the right and high-frequency shadow-masking effects on the entire surface. These mesoscopic details can be accurately replicated with **CTN** and **WardN** models. The **PhongN** preserves the mesoscopic details as well, but the details too more pronounced and partially baked into the diffuse term. Additionally, the appearance is darker towards the grazing angles in the left side, as it cannot model Fresnel reflectivity. The mesoscopic details for the **CT**, **Ward** and **Phong** models are significantly blurred, which can be problematic for finding pixel correspondences with optical flow. In the BTF reference image of the **felt** sample (right) we can see mostly diffused reflectance with large scale shadowing-masking and subsurface scattering effects. The effects cannot be replicated with either model, and the appearance is overall less structured. However, the **CTN**, **WardN** and **PhongN** models can preserved the higher frequency details.

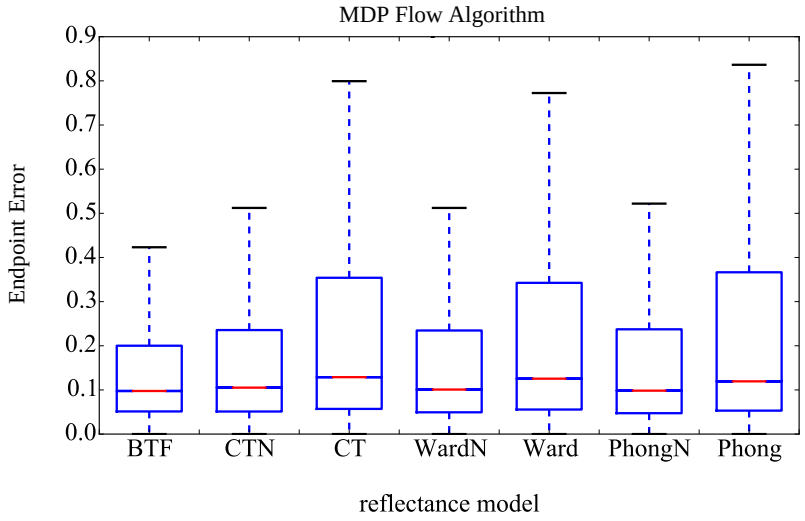

Figure 5.6.: Endpoint Error distributions of *MDP* Flow Algorithm combined for all 9 material samples and 39 image pairs. The BRDF models with normal estimation have similar median errors, yet slightly larger outliers. This indicates that the simulation quality, in general, matches the BTF model, but that some critical reflectance effects are more pronounced. BRDFs without normal estimation exhibit significantly increased median errors as well as outliers. Likely, the blurred geometric details cause the overall increased errors as optical flow algorithms depend on clearly defined structures. Differences between the BRDF models cannot be extracted and need to be assessed in a more detailed analysis.

els, combined for all nine materials and test sequences. For the *MDP* algorithm the median EPE is low and below $0.1px$, which corresponds to a relative flow error of slightly more than 1.3%. However, the distribution is heavy-tailed with mean and third quantile EPEs of $\approx 0.2px$, indicating that some reflectance effects are responsible for outliers in the flow fields. The distributions of the **CTN**, **WardN** and **PhongN** models are of similar shape compared to the BTF model and almost identical in direct comparison. The median EPE of the BRDF models is only slightly increased, but the mean and third quantile EPE are more abundant with $\approx 0.25px$, indicating that some simulated effects are more challenging for the *MDP* algorithm. By contrast, the **CT**, **Ward** and **Phong** models exibit larger differences in the EPE distributions. The median errors are increased by 50% to $\approx 0.15px$ and third quantile and mean errors are almost doubled. This indicates that the loss of high-frequency geometric details is critical to the *MDP* algorithm performance, as it relies on structures to find accurate pixel correspondences between frames.

H&S EPE Distributions For the *H&S* algorithm, the BTF errors are significantly more abundant in general. Compared to the *MDP* algorithm, the median errors are more than twice as large, and more importantly, the outlier errors are increased almost 10-fold. In fact, the outlier EPEs are almost as large as the average length of the ground truth flow vectors, indicating

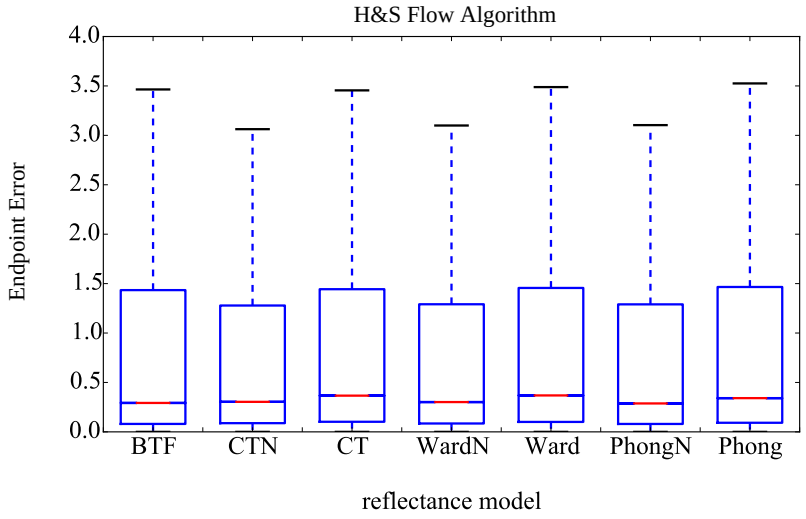

Figure 5.7.: Endpoint Error distributions of *H&S* Flow Algorithm combined for all 9 material samples and 39 image pairs. The EPE is significantly increased compared to the *MDP* algorithm. Outliers are in the range of the average length of the ground truth flow vectors. This indicates that some simulated effects pose a great challenge to the algorithm. As the overall performance of the flow method is poor, all EPE distributions are similar, and differences are not very pronounced. The BRDF models with normal estimation have almost identical median errors, compared to the BTF reference. However, outliers are slightly decreased. The median errors of the models without normal estimation are increased by $\approx$ 25%. This indicates that the blurred structures also affect *H&S* algorithm in its ability to find correspondences under seemingly unchallenging conditions. Again, there are hardly any differences between the BRDF models for the general flow performance.

that the algorithm can generally not deal with the displayed reflectance effects. The BRDF models with normal estimation have almost identical median and mean errors, but slightly decreased outliers compared to the BTF reference. This indicates that, in general, there are some minor differences in BRDF images for the reflectance effects that are most critical for the *H&S* method.

The models without normal estimation have overall larger errors, with a $\approx$ 25% increased median error. The increased median errors are likely caused by the loss of high-frequency details in the models. The overall more substantial errors of the models without normal estimation, offset the slightly decreased outlier EPE of the models with normal estimation back to the level of the BTF reference. However, as the performance of the *H&S* algorithm is rather poor, the differences between the reflectance models are, in general, not very pronounced. Compared to the *MDP* algorithm, the *H&S* algorithm seems to be more concerned with an overall plausible estimation of flow fields and less concerned with high-frequency mesoscopic details and structures.

Material Specific EPE Analysis The material specific EPE distributions can give a better indication of the most critical reflectance effects for optical flow performance analysis.

The almost diffuse carpet, wall and felt materials exhibit little mean and median errors below $0.1px$ for the *MDP* method and well below $0.2px$ for the *H&S* method. The third quantile EPE is also small, and outliers are small. As expected, the ma-

terials provide no challenge for flow algorithms, and all BRDF models perform virtually identical to the BTF method. This is an exciting finding, as differences in the appearance occur due to the lack of simulation of subsurface scattering effects (*c.f.*, Figure 5.5), which are seemingly are not critical to both used flow algorithms.

Stone and wood exhibit increased Fresnel reflectivity at low grazing angles, but are otherwise not very challenging. However, the stone and wood samples have only low to medium EPE values, with a median and mean EPE of $\approx 0.1px$ for the *MDP* method and $\approx 0.2px$ for the *H&S* method. In comparison to the diffuse material samples, the outliers are not significantly increased. By analyzing pixels at low grazing angles, I could identify no apparent increase in the EPE values. This shows that simulation of the Fresnel reflectivity is not critical to optical flow performance analysis for otherwise simple materials.

By contrast, the structured leather material sample with Fresnel reflectivity has slightly higher EPE values, with a median of $\approx 0.15px$ for the *MDP* method and $\approx 0.25px$ for the *H&S* method. The errors are virtually identical for the **WardN** and **CTN** models, but are lowered for Ward and **CT** and the *MDP* method and increased for the *H&S* method. The **PhongN** and **Phong** models exhibit the same behavior but have overall decreased error values. This matches the observation in Figure 5.5, that grazing effects are not simulated and the appearance is overall more defined. The leather results show that the differences in the geometric details lead to differences in the flow fields. Increased Fresnel reflectivity and anisotropic reflections seem to amplify the differences. The fabric material sample has overall smaller errors in the range of $0.1px$, like the other diffuse material samples. However, it exhibits similar saliences concerning the normal estimation and the BRDF model choice, as the leather sample. The only difference is that the **Phong** and **PhongN** errors are even more decreased and almost cut in half, likely due to the lack of anisotropic reflectance modeling.

As expected, the copper, aluminum samples have the highest errors and the substantial outliers in the BTF reference flow fields. The median EPE values are close to $0.2px$ for the *MDP* method and larger than $5px$ for the *H&S* method. This shows that specular highlights are the most critical to optical flow algorithms. In the case of the *H&S* method, the effects are so

problematic that unreasonable performance levels are achieved. Likely, the method almost always detects the motion of the moving highlights. For the *H&S* method, all BRDF models with and without normal estimation perform similarly poor. For the *MDP* method, the three BRDF models with normal estimation have a similar distribution, but overall slightly increased errors, indicating that the highlights are even more pronounced. The BRDF models without normal estimation have notably lower median errors of $\approx 0.15px$, but significantly increased outliers. This might be caused by slightly mislocated highlights and a lack of textures, as observed in Figure 5.5.

In the following sections, I will more thoroughly analyze the most prominent observations from the EPE distribution analysis, namely the highly specular metal samples and the structured leather and fabric samples with increased Fresnel reflectivity and anisotropic reflectance.

5.5.4. Reflectance Aware Specular Material Analysis

To study the effect of specular reflections in more detail, I analyzed the aluminum and copper results. Specular effects occur at small angles between the view vector(ω_o) and reflection vector(ω_r). I thus analyzed the specular material samples with respect to the reflectance aware EPE distributions around the perfect reflection direction (*c.f.*, Section 5.3).

Figure 5.8 depicts the reflectance aware EPE in the form of heatmaps. The heatmaps break down how the error sizes are distributed among the angle intervals, normalized per column by the total number of pixels which fall into the respective interval. In total, each heatmap shows the resulting errors of $\sim 15M$ pixels. The differences between the heatmaps of the BRDF models and the BTF model tell us how closely the models simulate critical specular effects for the flow algorithms.

For the *MDP* method of the BTF reference, one can observe that most errors fall into the interval $(0.06, 0.12]$. The absolute errors increase slightly for larger angles. The algorithm is mostly unaffected by the main specular highlights, possibly due to strong adaptive regularization. In fact, the largest errors occur for off-specular reflections, likely due to more substantial induced motion from the moving specular highlight and a loss of texture in the darker areas.

The **CTN**, **WardN** and **PhongN** distributions are almost

identical to BTF. However, the errors are overall slightly larger, most notably for angles above 12°. As observed in Figure 5.5 the highlight intensities of the BRDF model fits is higher due to the model based interpolation of the BTF samples. This likely causes the increased errors main specular direction for the *MDP* flow algorithm. The errors around off-specular reflections are possibly increased, due to the slightly darker appearance.

The **CT**, **Ward** and **Phong** have significantly larger errors for all angles and most prominently for angles above 6°. Likely this is caused by the 3° offset from the surface normals, which results in a 6° offset for the reflection direction.

For the *H&S* method, the BTF errors are overall more abundant. In contrast to the *MDP* method, one can observe higher errors for smaller angles. However, there appears to be no explicit dependency on the reflection vector as the algorithm performs poorly for all angles close to the main reflection. The *H&S* method seems to be in general, not capable of handling challenging highly reflective materials and gives no distinction between the reflectance model quality for optical flow performance analysis. Only minor differences lie in the off-specular reflections for angles $\geq 15°$. Here, the errors for the **CT**, **Ward** and **Phong** model are increased compared to the BTF reference. The **CTN**, **WardN** and **PhongN** model match the EPE distribution better. Likely the loss of structures causes the differences for the models without normal estimation.

I conclude that all models with normal estimation are well-suited for optical flow evaluation on specular materials and needed to preserve the rankings of the flow algorithms. There are some minor increases in the error values compared to the BTF model. However, as the BTF is not perfect for such highly reflective materials due to linear interpolation, this is a strong indication that BRDFs are more accurate for this material class. Using no normal estimation does change the distribution of errors. As a result, the models are less accurate for optical flow performance analysis of highly specular materials.

5.5.5. Reflectance Aware Structured Material Analysis

To study materials with fine geometric structures and anisotropic reflectance, I analyzed the fabric and leather samples. Similar to before I look at the reflectance aware EPE distributions with respect to the mirror direction. The specular highlights are less

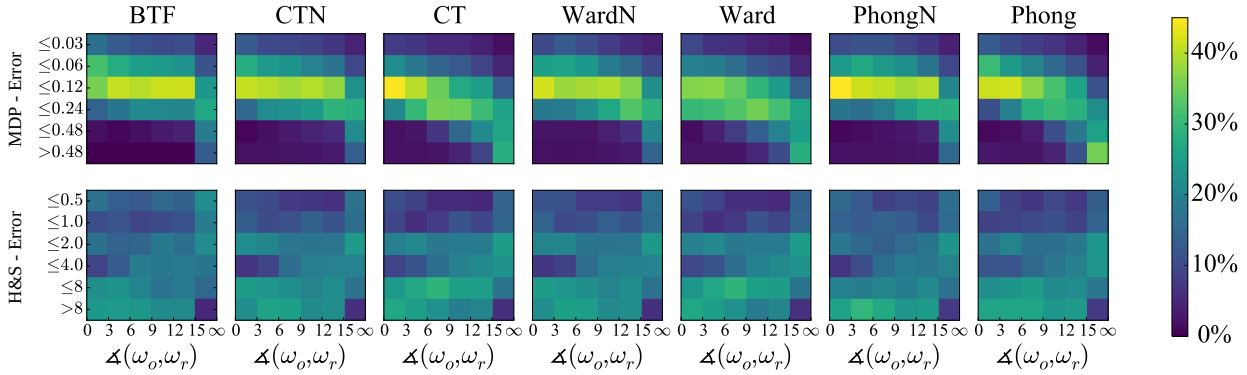

Figure 5.8.: Heatmaps show histograms of the reflectance aware EPE distributions for the specular aluminum and copper samples with respect to the angle between view and mirror direction. Color encodes relative occurrences. The *MDP* algorithm seems to be rather robust towards specular highlights, due to effective regularization. The highest errors do not occur around the main specular reflection, but for the off-specular reflections far from the perfect mirror direction. In those areas, the appearance of the images is darker and less structured, and the induced motion from the moving specular highlights is larger, overall harder conditions for flow algorithms. **CTN**, **WardN** and **PhongN** perform virtually identical. However, the errors are slightly larger for all angles, as the intensity of the highlights is increased due to a more accurate model-based interpolation of the BTF samples. The distributions of **CT**, **Ward** and **Phong** do however deviate due to the tilted reflection direction. The *H&S* method performs almost equally poor for all models, with most errors around the length of the ground truth flow vectors. It seems to be incapable of handling specular highlights regardless of the angle to the reflection direction. Small differences lie in the largest angle interval. Here, the errors for the models without normal estimation are larger due to a loss of mesoscopic structure. The models with normal estimation match the EPE distribution of the BTF model.

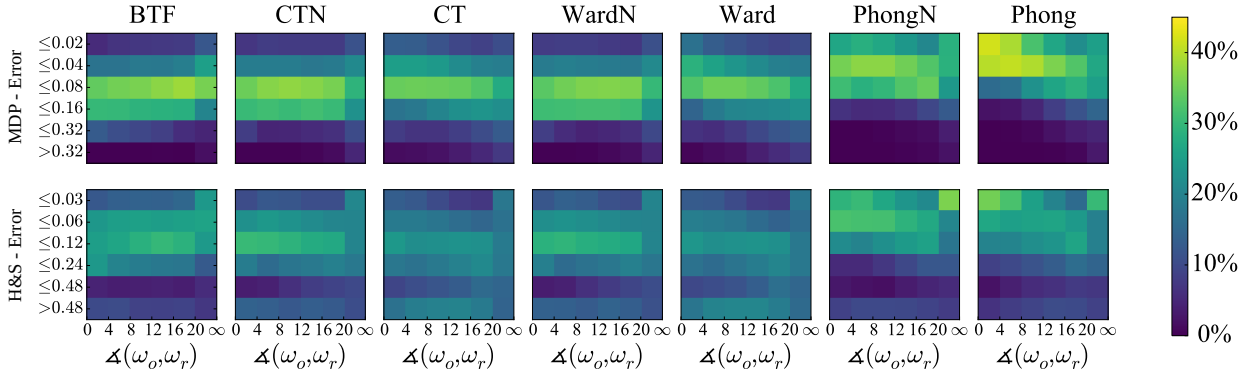

Figure 5.9.: Heatmaps of the fabric and leather samples as in Figure 5.8. Both flow algorithms perform reasonably well for the contemplated materials. The errors of the BTF reference are largest around the reflection direction and decrease for off-specular reflections. This can be replicated with the **CTN** and **WardN** models and both flow algorithms. The flow performance with the **Ward** and **CT** models are ambiguous. For the *MDP* method, the errors are decreased towards the reflection direction and increased for larger angles. This is reversed for the *H&S* algorithm. The blurred mesoscopic details are the likeliest cause and seemingly affect the algorithms differently. Therefore, normal estimation is needed to closely match the BTF reference and to preserve the rankings of the algorithms. The **Phong** and **PhongN** models perform poorly, with overall decreased errors for all angles. As the model cannot capture anisotropic reflectance, it is baked into the diffuse reflectance texture and thus motion-invariant. Those are non-discriminating conditions for both flow algorithms.

intense and broader. Thus I binned the errors to larger angle intervals in the error heatmaps from Figure 5.9.

In contrast to the specular materials, the *H&S* method performs reasonably well for the leather and fabric samples. For the BTF Reference and the *MDP* algorithm most errors fall into the $(0.04, 0.08]$ and $(0.08, 0.16]$ intervals for all angles close to the mirror direction. For the *H&S* method the errors are overall $\approx$ 50% increased and mostly fall in the $(0.06, 0.12]$ and $(0.12, 0.24]$ intervals. However, the EPE distributions are broader and distributed with a higher variance. In contrast to the specular materials, the EPE are slightly decreased for angles above 20°. For those angles, the variations of the diffuse reflectance and the structured geometry lead to easier pixel correspondences for the flow algorithms.

CTN and **WardN** exhibit almost identical distributions as BTF for both algorithms. However, the errors are slightly smaller close to the reflection direction and slightly larger for off-specular directions. This is possibly caused by sharper main highlights and a slight loss of structure towards the off-specular directions. For the **CT** and **Ward** models without normal estimation this effect is severly amplified for the *MDP* algorithm. The errors close to the reflection direction are significantly reduced, and the errors in off-specular direction are severely increased, likely caused by the blurred mesoscopic details. By contrast, the effect is reversed for the *H&S* algorithms. The errors are increased for small angles and barely changed for far angles. Likely, the algorithm is more affected by the loss of structure for model fits without normal estimation.

The **PhongN** and **Phong** models perform poorly. The errors are significantly reduced overall, likely due to the non-existent anisotropic reflectance modeling. As a result, most of the reflectance is encoded in the motion independent diffuse term, creating easy conditions for the flow algorithms.

I conclude that the **Phong** model is insufficient for complex and anisotropic materials regardless of normal estimation. The **CT** and **Ward** models simulate the BTF model closely over all angles. Normal estimation further improves the simulation and is needed not to affect the ranking of the algorithms.

5.6. Summary and Outlook

In this chapter, three analytical BRDF model were studied in the context of optical flow dataset generation: A computationally inexpensive Blinn-Phong model, an anisotropic Ward model, and an anisotropic physically based Cook-Torrance model. In contrast to the proven bidirectional texture functions, they are more efficient, easier to process, supported by conventional rendering systems and can better model sharp specular reflections. However, they lack the capability to model subsurface scattering and large scale shadowing and masking effects.

The models were fitted to various material samples of the BTF database Bonn[156], using novel approaches for spatially consistent BRDF model fits and a normal estimation method to counter parallax artifacts and the inaccuracies of the coarse BTF geometry on the mesoscopic texel scale. The normal estimation could increase the accuracy of the fits by 25% on average and more than 50% for structured materials with high-frequency geometric details. For all materials, a Cook-Torrance fit was the most accurate with a small margin to the Ward model and significant a margin to the Phong model.

Finally, optical flow experiments were conducted on unbiased renderings of randomized scenes with the reflectance models applied. To assess the most critical optical flow effects and the BRDF simulation accuracy, I analyzed the perceptual appearance of the synthetic images and the computed flow fields with a reflectance aware Endpoint Error distribution.

I showed that subsurface scattering and large scale shadowing and masking effects are mostly irrelevant to optical flow performance analysis. The effects create perceptual differences between the BTF and BRDF model renderings. However, as the differences are diffused and mostly motion-invariant, they do not affect the computed flow fields.

For mostly diffuse and isotropic materials, the choice of the BRDF model hardly matters for optical flow algorithms. Even the simplest model, a Blinn-Phong fit, is as good as the computationally expensive reflectance field rendering. This changes when the reflectance of the materials is anisotropic. The specular Phong term cannot model the anisotropic effects and are baked into the diffuse term. As a result, the reflectance is structured and motion-invariant, which are unchallenging conditions for optical flow algorithms. An anisotropic Ward or Cook Tor-

rance model can, in this case, be used to match the results of the BTF model closely.

For structured materials with fine geometric details, the BRDF models without normal estimation perform poorly. The structures are diffused, and small highlights are missed. As a result, the optical flow algorithms perform vastly different than for the BTF reference data. The normal estimation can then be used to create BRDF models with almost perfect accuracy. However, the Phong model with normal estimation is a worse choice than the more sophisticated Cook-Torrance and Ward models, as structures are too pronounced and specular details are lost.

As expected, metallic materials with highly motion-invariant specular highlights are the biggest challenge for optical flow algorithms. BRDF model fits with normal estimation can be used with high accuracy, that can even exceed the quality of the BTF model, which is limited for mirror-like reflections. The choice of the BRDF model does hardly affect the algorithms. Only for off-specular reflections far from the highlights the Cook-Torrance and Ward models are slightly superior to the Phong model.

We now better understand which BRDF model fulfills the requirements for optical flow performance analysis and different material types. BTFs created from reflectance field measurements are highly accurate and only limited for mirror-like reflections. The best performing BRDF model, a Cook-Torrance model with normal estimation, performs virtually identical for optical flow performance analysis and arbitrary material types. For the most challenging highly reflective surfaces, the model has even the potential to outperform the BTFs.

By reprojecting, the raw reflectance field data to the coarsely scanned surface geometry information is lost. This can be compensated with normal estimation for delicate mesoscopic, but not for macroscopic geometric details. By skipping the reprojection and estimating the geometry directly from the reflectance field data, the perceptual differences between the synthesized images can be further minimized even for large scale shadow-masking effects. By resampling and compressing the reflectance field data into it into the BTF format, angular and spatial reflectance information is lost, which cannot always be restored by the BRDF models. Therefore, it would be beneficial to skip the intermediate BTF representation, overcoming the limitations for highly reflective materials.

A further improvement could be the adoption of the GGX distribution, which is used by the widely adopted Disney PBR model [15]. This distribution is heavy-tailed, has been shown to match measured data better [152], and could therefore potentially create satisfying result with even a single specular lobe.

6

Augmentation of Physically Plausible Precipitation Effects

6.1. Motivation

Deep neural networks (DNNs) are trained on a limited amount of training data to generalize to arbitrary testing conditions, *i.e.*, the network should perform adequately and robust under unseen conditions. The straight forward approach to improve network generalization is to utilize an increased amount of training data displaying a large variety of scene conditions. However, this approach is inherently expensive and limited, as it is hard to ensure that the training set sufficiently covers all critical conditions.

As an alternative, data augmentation can be used to expand the training dataset and increase the robustness of the networks towards critical and unseen conditions, *c.f.*, Section3.4. Spatial transformations are incorporated in most training schedules to ensure that the training data is not overfitted. Other effects that can be modeled phenomenologically, *e.g.*, noise, blur, are also often applied for various computer vision tasks to increase network robustness [141, 56, 146, 95]. However, data augmentation can only mimic effects in image space and effects that originate from the scene, *e.g.*, reflections, cannot be modeled.

One example are precipitation effects. They are challenging for many automotive computer vision tasks, due to reflections from puddles and road surfaces, blurring from droplets on the windshield, and lowered contrast from water spray and mists. However, they heavily rely on scene context, *e.g.*, the appearance of wet surfaces depends on the camera position, scene lighting, material properties, and most importantly, the surface geometry.

On the other hand, many datasets include accurate scene geometry and camera poses, screen-space techniques can trace reflections of a 2D image with depth information [99], and the literature provides reflectance and scattering models for wet surfaces and water particles (*c.f.*, Section 2.5.2). This led to the idea of

Figure 6.1.: I present a technique for augmenting precipitation effects in automotive datasets. The method was implemented for the HD1K Benchmark Suite[1] dataset. (a) depicts an example image of that dataset. The original image data is augmented with physically based ground surface wetting, including the water saturation of the surface and the accumulation of puddles. In (b) the original image is augmented with a low amount of ground surface wetness, where water has not yet accumulated. In (c) the wetness level is increased, and puddles start to form. In (d) the full precipitation augmentation model is applied. This includes the maximum ground surface wetting, water particles from water spray and mists, as well as droplets on the windshield.

combining these data and models to create a novel technique for physically plausible precipitation augmentation.

The method, presented in the following chapter, was implemented for the HCI Benchmark automotive dataset, due to availability of accurate scene geometry and camera poses [72, 71].

6.2. Contributions

In the following, I present a method for the augmentation of automotive computer vision datasets with precipitation effects. Figure 6.1 depicts examples of the produced images. The method utilizes physically based models for wet surfaces and light scattering and creates consistent and plausible results by working in 3D world space. The main contributions, presented in this chapter, shall be subsequently summarized.

A consistent virtual scene can be extracted from a computer vision automotive dataset with point cloud and camera pose data. I present parametrized models of ground surface water levels, water spray and mists, and droplets on the windshield to enrich the scene with precipitation effects. The procedural water

level distribution respects height and ground material types. As a result, water plausibly accumulates at local minima of the surface, *e.g.*, next to curbs and in ditches, and for surfaces where water seeps slowly, *e.g.*, the road. The droplets blend, merge and stick to the windshield and the spray and mist distribution respect ground level distance.

This consistent virtual scene description then serves as the baseline for the proposed augmentation. The original images from the clean dataset are UV-mapped onto the virtual scene, by perspective projection from the provided camera poses. Subsequently, I approximate material properties, *i.e.*, surface normals, surface roughness, and surface porosity, by extracting spatial variations from the original image data and referencing the materials to measurements of urban ground surfaces. I then show, how this information can be used to separate diffuse and specular image intensities by performing a screen-space specular light transport approximation. The same method can be used to compute novel specular and scattering light transport for augmented scene elements. Finally, this enables a plausible augmentation of various precipitation effects; 1) The diffuse darkening and specular boost of water penetrated ground surfaces. 2) The reflections of accumulated water in drenched regions and puddles. 3) The scattering and absorption of water spray and mists. 4) The scattering and blurring from water droplets on the windshield.

6.3. Creating Virtual Scenes

A virtual 3D scene representation of the original automotive sequence is needed for the augmentation of precipitation effects. The ground truth depth map could be used as scene geometry, but a consistent and structured scene representation is better suited to ensure spatiotemporal stability of the augmentation. I commence with the reconstruction of structured meshes from the measured point clouds of the static environment. Next, I show how the geometry of dynamic objects can be approximated with cardboards. I then detail my method for creating plausible water level distributions that respect ground surface height and material properties. Finally, I present my proposed models for water droplets on the windshield and water spray and mists.

The method was implemented for the HCI automotive dataset [72, 71] which was published as the HD1K Benchmark Suite-

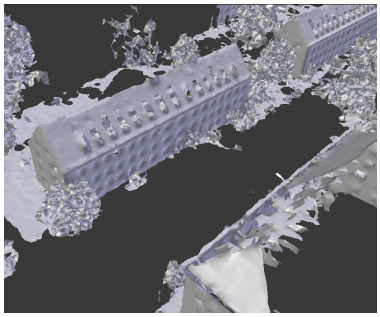

Figure 6.2.: The background environment is reconstructed with a coarse mesh. Artifacts occur for non-convex objects, such as the vegetation. However, the background environment is only augmenting the original image via reflections and, therefore, inaccuracies can be tolerated.

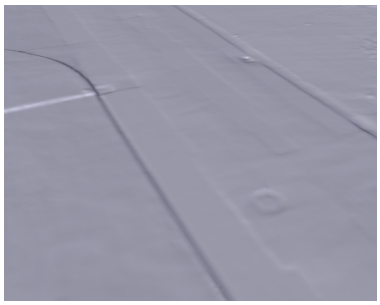

Figure 6.3.: The ground surface mesh is reconstructed with a finer mesh (5M triangles). Since the surface is mostly flat, it can be reconstructed with high accuracy. Curbs, road ditches, sewer lids, and even small cracks and variations in the road surface are faithfully reconstructed. Such high-fidelity detail is beneficial to my method, as the wetness model directly modifies the ground surface.

for optical flow. The dataset is well suited to create a consistent virtual 3D scene representation, due to the availability of a high-fidelity point cloud of the static environment and accurate camera poses. However, the method is dataset independent and could be applied to any dataset that features point cloud data, camera poses, and dynamic object or segmentation masks.

6.3.1. Mesh Reconstruction of the Static World

The flow and depth ground truth of the underlying HCI automotive dataset was assembled with point cloud measurements from a survey grade RIEGL VMX-250-CS6 laser scanner system [72]. The point clouds are dense and capture static scene geometry with an accuracy of $\approx 1cm$. However, such unstructured point cloud data cannot be directly used for the proposed method. The light transport simulation, the height aware water levels, and the material reflectance model all require continuous surface normals and a structured mesh representation of the scene.

To create suitable meshes from the facilitated point data, the open-source mesh processing tool "Meshlab" was used [18]. Surface reconstruction from unstructured point clouds rarely leads to good results for non-convex objects. The point relations are often ambiguous, and points are connected based on proximity even though they have no link in the real-world, *e.g.*, consider a tree where the point data of neighboring branches should not be connected. Typically, only the background environment, *i.e.*, buildings and vegetation, contain such problematic areas. However, the appearance of the environment is not augmented and only visible via reflections. Small geometric inaccuracies are therefore acceptable and can still lead to plausible results. By contrast, the ground surfaces are the ones being augmented. However, they are mostly flat and, thus, more easily meshed. To ensure the best possible quality of the ground mesh reconstruction, I separated the ground level point set from the remaining point cloud data and meshed both point sets separately.

To this end, a screened Poisson surface reconstruction was applied, which produces excellent results for unstructured point sets [69]. As the method requires accurate normals for each data point, they were estimated with the `computeNormalsForPointSet` function for 20 point neighbors and w.r.t. a view direction along the positive z-axis, ensuring that normals are not pointing down-

wards. I then meshed the point sets with an octree depth of 13 for the ground plane and an octree depth of 12 for the environment. This resulted structured meshes with roughly 5M and 2M triangles, respectively (*c.f.*, Figure 6.2 and 6.3) .

6.3.2. Cardboard Meshes of Dynamic Objects

To simulate plausible reflections of the dynamic elements, *e.g.*, cars and pedestrians, some form of dynamic geometry is needed. However, the utilized dataset only includes point cloud data for the static environment and lacks geometry measurements for the dynamic objects of the scenes. To approximate the geometry of dynamic objects, Honauer proposes to place 2D cardboards in the virtual world by projecting the dynamic object masks with the ground truth depth of the lowest mask point in image space [60]. These cardboards are always aligned perpendicular to the view direction and work reasonably well for pedestrians in typical walking poses and at least one foot on the ground. For vehicles, I adjusted the previous approach by fitting a 2D cardboard through two local minima of the dynamic masks, which, in most cases, correspond to the two lowest wheel positions on the ground. The resulting vehicle cardboards can, therefore, be rotated in arbitrary directions around the z-axis in world space. While these cardboards are certainly not an exact representation of the vehicle geometry, they are an improvement over the conventional cardboard approach and better suited to deliver plausible reflections.

6.3.3. Masked and Height-Correlated Procedural Water Distribution

This section is concerned with the process of creating plausible water level distributions for a given global wetness parameter of the augmentation model. As described in Section 2.5, the wetting process of porous dielectric materials, *i.e.*, the type of materials that make up urban ground surfaces, is divided into two phenomena: First, the saturation and water penetration of the pores, which causes a diffuse darkening and slight increase in specular reflectivity. Second, the accumulation of water, which can no longer penetrate the pores and instead forms drenched regions and puddles.

We are, therefore, looking for a plausible water level distribution W for each point on the ground surface and a given global

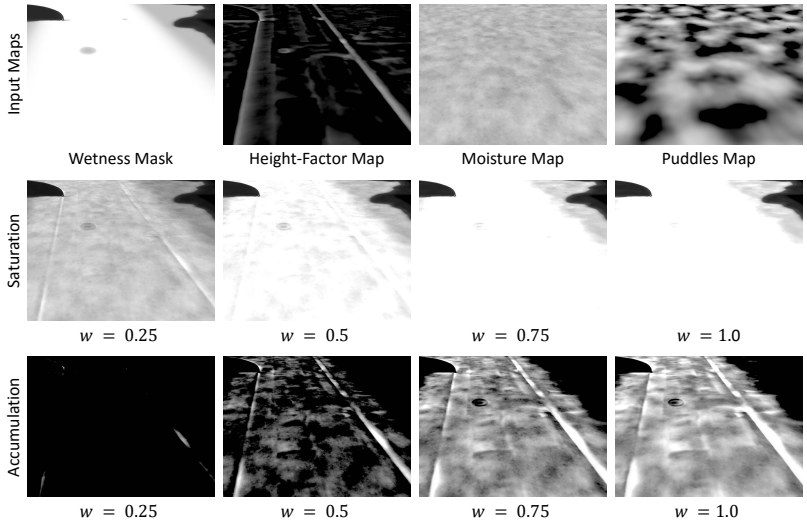

Figure 6.4.: Four texture maps are combined with the proposed water distribution model. The manually created *wetness mask* acts a multiplier of the model. Surfaces where water can fully accumulate, *e.g.*, the road, are marked in white, surfaces which should not be affected, *e.g.*, vegetation, are marked in darker colors. The *height-factor map* was retrieved from local minima of the ground surface height map. White marks regions where water likely accumulates due to gravity, *e.g.*, the ditches on the roadsides. The *moisture map* is a high-frequency procedural texture function for mimicking wetness variations due to material properties and ground surface drying. The *puddles map* is a low-frequency procedural texture function for mimicking puddles and drenched regions. The middle row images depict the ground surface water saturation for various global wetness levels w of the augmentation model. The bottom row depicts the corresponding water accumulation.

wetness parameter:

$$W(u, v, w) \in [0, 1] \tag{6.1}$$

where u, v are the ground plane UV-map texture coordinates which are orthographically projected from the top view with the unit m, and w is the global wetness parameter of the augmentation model. Subsequently, a water level $W(u, v, w) = 0.5$ refers to full saturation of the ground surface, and a water level $W(u, v, w) = 1.0$ refers to an accumulation of water that is large enough to cover the ground surface fully.

Precise modeling of W requires knowledge about physical material properties, *e.g.*, the porosity and internal seepage, as well as scene context, *e.g.*, the local height profiles and drains. Out of those, only the local height information can be extracted from the dataset. Nevertheless, perceptional plausible distributions of water saturation and accumulation can be created. To this end, I propose to combine a manually created wetness mask, extracted height information from the dataset, and two procedural texture maps for wetting: First, a procedural moisture map with high-frequency details, which are typically present due to variations of porosity and internal seepage. Second, a procedural puddle map with low-frequency details for the accumulation of puddles. The maps and visualizations of how they form the water distribution models of saturation and accumulation for different wetness levels w are depicted in Figure 6.4. Subsequently, I describe the synthesis approach of each map and how they are combined to formulate the proposed model f W.

Wetness Mask

Not all ground surface are equally affected by wetness and, therefore, some surfaces are more likely to form drenched regions and puddles than others. To accommodate for material properties, I manually drew a wetness mask $M(u, v)$ on the ground mesh, acting as a multiplier for water accumulation. Three classes of ground surfaces are masked: First, ground surfaces which are hardly affected by wetness, *e.g.*, vegetation. To this class, I applied $M(u, v) = 0.25$ so that water penetration of the surface cannot fully apply. Second, ground surfaces where water seeps rapidly and can hardly accumulate, *e.g.*, soil and drains. I chose $M(u, v) = 0.5$ for this class so that the total wetness can never exceed full penetration levels. Third, ground surfaces where water seeps slowly and can quickly accumulate, *e.g.*, road and sidewalks. For this class the wetness mask is not attenuated with $M(u, v) = 1.0$ and water levels can reach full accumulation. Subsequently, the mask was blurred to avoid sharp and unrealistic transitions. While the process of drawing a material mask is rather tedious, it quickly provides good results for a small scale dataset, *e.g.*, the HD1K dataset used in this study. For a large scale dataset, manually masking is not feasible. However, a reasonable wetness mask could be derived from segmentation ground truth.

Height-Factor Map

Water gathers around local minima of the ground level, due to gravity. With increasing wetness, the extent of the accumulation grows, and ground level minima at larger scales become relevant. To model the influence of the ground level, I extracted the local minima of the ground surface height map with difference of Gaussians filtering. The size of the Gaussian kernels is thereby correlated with the wetness level to model local minima on different scales:

$$F(u, v, w, k) = 1 + k * (G(u, v, \sigma_w) \circledast H(u, v) - H(u, v)) \quad (6.2)$$

where $H(u, v)$ is the ground surface height map, $G(u, v, \sigma_w)$ is a Gaussian Kernel with standard deviation $\sigma_w = w/2$, and k is a weighting factor which parameterizes the influence of the height differences. The height map H, and the standard deviation σ_w

have the unit m. The weighting factor k has the unit m^{-1} and is chosen empirically for each procedural wetness map. I chose a factor of $k_{acc} = 100m^{-1}$ for the water accumulation map, *i.e.*, the accumulation is doubled for a height difference of $0.01m$. I further chose a factor of $k_{sat} = 2m^{-1}$ for the water saturation map, *i.e.*, the ground surface water saturation is doubled for a height difference of $0.5m$.

Procedural Wetness Map

The wetness mask models the influence of material properties on the water distributions at a large scale, the height factor map models the influence of the ground surface height on medium to large scales. The resulting water distribution from those two maps alone are rather homogeneous. However, realistic water distributions are typically varied on smaller scales, *e.g.*, from spatial variations of material properties or dynamic forces. I propose to model the remaining factor with procedural maps to create perceptional plausible variations of the water distributions.

Ground surface water saturation depends on high-frequency details, *e.g.*, from variations of the internal seepage or the progress of the drying process. I created a water saturation map with the procedural moisture pattern from the procedural texture modeling tool Substance Designer [2]. I used the maximum moisture pattern scale of eight with a resolution of 8192×8192 texels, which is subsequently mapped to repeating areas of $40m \times 40m$. The procedural pattern is then combined with the height factor to form the water saturation model:

$$S(u, v, w) = lerp(S_0(u, v), 0.5, 0.75 * w) * F(u, v, w, k_{sat}) \quad (6.3)$$

where $lerp()$ is the linear interpolation function with the third parameter as the blending weight, $S_0(u, v)$ is the moisture pattern, and $F(u, v, w, k_{sat})$ is the height factor as in Equation 6.2 with the weight factor $k_{sat} = 2m^{-1}$. The linearly blending of the pattern with 0.5 reduces the contrast of for higher wetness levels and is motivated by the observation that for higher wetness levels the water saturation becomes more homogeneous.

[2] www.substance3d.com

Puddles and drenched regions are typically formed at small to medium scales. I created a water accumulation map with procedural $1/f^4$ Perlin noise [54], were the pattern size of the largest scale was set to two meters. The noise pattern was subsequently contrast enhanced with a factor of four. Additionally, I applied a brightness offset of -0.25 intensity levels, to model a maximum puddle coverage of around $1/3$ of the ground surface area. The procedural Perlin noise is then combined with the height factor to form the water accumulation model:

$$A(u, v, w) = lerp(A_0(u, v), 0.5, 0.5 * w) * F(u, v, w, k_{acc}) \quad (6.4)$$

where $lerp()$ is the linear interpolation function with the third parameter as the blending weight, $A_0(u, v)$ is the contrast enhanced Perlin noise, and $F(u, v, w, k_{acc})$ is the height factor as in equation 6.2 with the weight factor $k_{acc} = 100m^{-1}$. Again, the linearly blending of the noise function with 0.5 is chosen to create a more homogeneous water distribution for higher wetness levels.

Proposed Composite Distribution Model

The proposed model for a perceptionally plausible water distribution is a composition of the previously presented texture maps, which depends on the global wetness level:

$$W(u, v, w) = w * M(u, v) * lerp(S(u, v, w), A(u, v, w), 0.5 * w) \quad (6.5)$$

where $lerp()$ is the linear interpolation function with the third parameter as the blending weight, $M(u, v)$ is the manually drawn wetness mask as in Section 6.3.3, $S(u, v, w)$ is the water saturation model as in equation 6.3, and $A(u, v, w)$ is the water accumulation model as in equation 6.4. The linear blending of the water distribution models is motivated by the observation that initially the ground surface is penetrated by water before it can accumulate to drenched regions and puddles.

6.3.4. Water Spray Distribution Model

In rainy scenarios, water spray and mist, *i.e.*, a suspension of small water particles and dirt, are often present in the air close

to the ground level Typically, the water spray is swirled up by dynamic motions, such as from vehicles tires and to a smaller extent from wind forces. Mists are substantially similar in terms of appearance but caused by the rapid cooling of water saturated air. Precise physical modeling of both effects requires measurements of environmental parameters, *e.g.*, vehicle geometry, vehicles dynamics, temperature, and wind. These parameters are not available. Instead, I propose to model the distributions of the mist and water spray phenomenologically. Ignoring external forces, the density of fog and mist decreases exponentially with the distance from the ground plane [57]. This can be modeled with a height based exponential gradient volume. For increased realism, I propose to further model external forces and environmental factors, which influence the density of the fog, employing the $1/f^3$ Perlin volume noise function with a base scale of $0.5m$ [54]. Both models are then multiplied to form the proposed density distribution model for water spray and mist augmentation:

$$M(x, y, z, m) = \frac{m}{50} * P(x, y, z) * exp(-d) \tag{6.6}$$

where m is the global water spray and mists parameter from the augmentation model, $P(x, y, z)$ is the volumetric Perlin noise function, and d is the minimum distance between x, y, z and the ground surface plane. The distribution model is scaled with the factor $\frac{1}{50}$ to provide an average view distance of $50m$ at ground surface height for $m = 1$. This corresponds to an average view distance of around $\approx 150m$ at realistic automotive camera height levels.

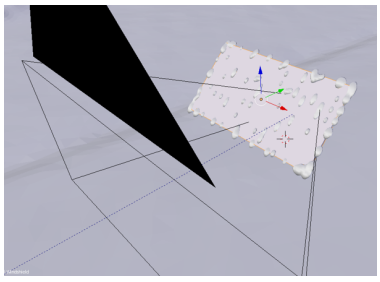

Figure 6.5.: A virtual windshield is placed in front of the virtual camera in a distance of $0.2m$ and an inclination angle of 45°. Metballs are then randomly distributed on the windshield's surface to simulate water droplets. These metaballs merge, blend, and stick to the windshield just like in the real-world.

6.3.5. Droplets on the Windshield

Water droplets on the windshield can obscure the vision of the camera. To simulate this effect, I placed a virtual windshield $0.2m$ in front of the camera with an inclination angle of 45°. While automotive cameras are in reality placed closer to the windshield, *i.e.*, the distance to the camera lens is typically only a few millimeters, this setup more closely resembles the testing setup of the original HD1k dataset. The optical properties of the real-world windshield are already included in the original dataset. Hence, I applied a transparent material to the windshield, not using it to augment the image directly. However, it acts as a geometric proxy for the water droplets, which are mod-

eled as blender metaballs and randomly placed on the windshield geometry. Metaballs are an implicit and procedural description of geometry, which was tailored explicitly for fluids [12]. The geometry is explicitly formed based on proximity of the metaballs and thresholds for cohesion and adhesion forces. As a result, the water droplets stick to the virtual windshield, and they merge and blend (*c.f.*, Figure 6.5).

6.4. Augmentation of Real-World Appearance

In this section, I present the augmentation of physically based wet ground surfaces, droplets on the windshield, and water spray and mists for real-world automotive datasets. In Summary, the augmentation consists of six steps:

- The input image data is projected onto the virtual scene of Section 6.3 and a virtual sky sphere for the background.
- Generic emission values are applied to scene surfaces and objects, which are either occluded or outside of the camera frustrum. The emission value of each object class is chosen as the mean object intensity of the original dataset. For the background sky sphere, the generic emission is set to 0.0 at ground level, and linearly increases to 1.0 for elevation angles above 15°.
- Specular BRDF parameters for the ground plane pixels are approximated from the input image data and projected onto the ground plane mesh.
- The diffuse and specular ground plane appearance is separated, utilizing the approximated BRDF parameters and a specular light transport approximation in screen-space.
- The previously separated appearances of the ground plane surfaces are augmented, utilizing the water level distribution model of Section 6.3.3 and a physically based wetness model which is closely related to model in Section 2.5.2.
- The scene is augmented with water particles, utilizing with the distribution models from Sections 6.3.4, 6.3.5 and a physically based BSDF model [15].

The method was implemented with open-source modeling tool Blender [11] and uses its path tracing engine Cycles for render-

ing. In the following, I detail each step of the augmentation method.

6.4.1. Projection of Real-World Appearance

To plausibly augment the image intensities in 3D space they need to be initially applied to the virtual scene from Section 6.3, *i.e.*, the static world mesh, the dynamic object cardboards.

To this end, I used a Blender thin-lens camera as a virtual camera representation of the original HD1K Benchmark Suite dataset. The thin-lens model is a reasonable approximation of the real-world camera, as the distortion of images is rectified, and the remaining intrinsic camera parameters are known. I set the camera pose, focal length, sensor resolution, and sensor-shift of the virtual camera model to accord to the real-world camera setup. In most cases, I used an aperture of $0.0m$ as the original image data is already affected by the aperture of the real-world optics. However, a different aperture can be set to simulate additional depth-of-field effects in 3D space realistically, *e.g.*, the depth-of-field blurring of water droplets on the windshield.

Subsequently, I mapped the original 2D image data with the Blender function `bpy.ops.uv.project_from_view`. The function performs a perspective camera projection, respecting the extrinsic and intrinsic virtual camera parameters. Additionally, I set flag `clip_to_bounds=True` to ensure that the image is not repeated outside of camera frustrum. I applied this procedure to all objects in the scene, including a virtual sky sphere background for distant objects and the sky lighting. Subsequently, I set up each object to emit the mapped intensity of the original image data. Tracing the virtual camera then results in a rendered image that is virtually identical to the original image data.

6.4.2. Modeling of Ground Plane BRDF Parameters

The proposed method uses the principled shading model from Disney [15] to augment physically based wetness effects on the ground plane surfaces. Specular BRDF parameter, *i.e.*, the index of refraction (IOR), the specular roughness, and shading normals, are needed for two subsequent stages: First, for the separation of the diffuse and specular ground surface appearance. Second, as the specular base parametrization for the ground surface wetness model.

Reference IOR and roughness values, extracted from literature and previous measurements, can be applied as spatially constant BRDF parameters. The reconstructed geometry normal can be used for shading. However, this leads to a smooth and homogeneous appearance of the ground plane surface. By contrast, real-world surfaces exhibit high-frequency variations of roughness and shading normals, *c.f.*, Chapter 5. The IOR value, on the other hand, is a material specific property and typically hardly varies in space.

Calibrated illumination and camera setups, capturing the appearance for a vast amount of view and lighting directions, are needed to measure BRDF parameters accurately (*c.f.*, Section 2.4.4 and Chapter 5). By contrast, the captured images of automotive dataset sequences do not meet the requirements for an accurate BRDF estimation: 1) The extrinsic parameters and radiometric intensities of the scene lighting are unknown. 2) The camera poses are not perfectly calibrated and often too similar. 3) The appearance of the surfaces is inconsistent, due to dynamic changes in scene lighting and potentially automated exposure control. Additionally, fitting BRDF models to a large amount of measured data is time-consuming and naturally contradicts the computational requirements of image augmentation for massive training datasets.

Consequently, I ruled out a full BRDF estimation from multiple captures. Instead, I modeled high-frequency variations of the BRDF parameters with filter-based approximations from the single input images. Subsequently, I modified reference values from literature and previous measurements. Similar approaches are often used for texturing materials in real-time computer graphics when only a single photograph is available. In the following, I outline the proposed approximations for three specular BRDF parameters.

Index of Refraction In many applications, especially for real-time rendering, the IOR is assumed constant and approximated with $n = 1.5$ for dielectric surfaces [15]. This IOR approximation is accurate for medium to high roughness values, even when the true IOR deviates. Applying a wetness model to the ground plane surfaces does, however, cause very low roughness values and sharp specular reflections. As a result, the $n = 1.5$ approximation becomes inaccurate for urban ground materials, which typically have higher IOR values. I propose to utilize $n = 1.63$,

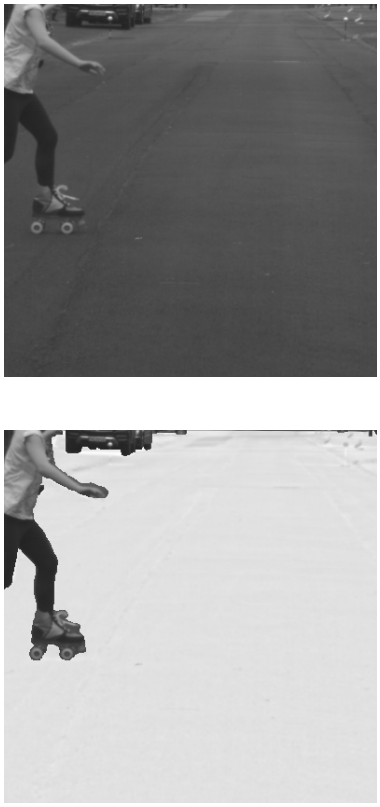

Figure 6.6.: **Top:** Original image for reference. **Bottom:** Same image where the appearance of the ground surface has been replaced with the estimated roughness parameter. Black depicts mirror-like specular reflectivity, white depicts diffused specular reflectivity. The roughness is estimated by difference Gaussians filtering and mapping the results to reasonable parameter bounds from measurements.

the IOR of tarmac, as the base IOR of the dry ground plane surface.

Roughness The specular roughness determines the overall strength of specular highlights. Assuming a mostly diffuse illumination, it creates high- to medium-frequency variations of the image intensities. Existing software tools for creating material textures from a single photograph, *e.g.*, the commercial Bitmap2Material [3] software toolkit, approximate surface roughness from high- to medium-frequency image features. I propose to extend this idea by additionally mapping the image features to physically plausible parameter ranges of urban ground materials.

I extracted the variations of the image intensities using difference of Gaussians filtering. To ensure, that dynamic foreground objects are not affecting the image features for the ground surface, I applied a masked filtering with the dynamic objects masks:

$$\Gamma_{\alpha,\sigma_0,\sigma_1}(x,y) = (G(x,y,\sigma_1) \circledast I(x,y)) - (G(x,y,\sigma_0) \circledast I(x,y)) \tag{6.7}$$

where $I(x,y)$ is the original image data and $G(x,y,\sigma_1)$, $G(x,y,\sigma_0)$ are masked Gaussian Kernels. For the ground surface roughness, standard deviations of $\sigma_1 = 8.0, \sigma_0 = 1.0$ led to good perceptional results. I then mapped the normalized feature map to realistic parameter ranges of urban materials, creating the final approximation of the ground surface roughness:

$$R(x,y) = lerp(\alpha_{min}, \alpha_{max}, \bar{\Gamma}_{\alpha,\sigma_0,\sigma_1}(x,y)) \tag{6.8}$$

where $lerp()$ is the linear interpolation function with the third parameter as the blending weight, $\bar{\Gamma}_{\sigma_0,\sigma_1}(x,y)$ is the difference of Gaussians image as in equation 6.7, which was normalized to the $[0,1]$ intervall, and $\alpha_{min}, \alpha_{max}$ are the lowest and highest roughness values. For the urban ground surface a choice of $\alpha_{min} = 0.5, \alpha_{max} = 1.0$ leads to visually plausible results.

Shading Normals The surface normals determine the main reflection direction of the incident light. Assuming mostly diffuse illumination, they create high-frequency variations in the image intensity, due shadowing-masking effects as well as high-frequency specular reflections, *c.f.*, Chapter 4 and Chapter 5.

[3] www.substance3d.com

Similar to surface roughness, the surface normals are extracted from highest-frequency image features by commercial software tools for single shot material texturing. Again, I propose to extend this idea by mapping the image features to physically plausible height variations of urban ground surfaces.

As for the surface roughness, I model the height variations with masked difference of Gaussians filtering:

$$\Gamma_{n,\sigma_1,\sigma_0}(x,y) = (G(x,y,\sigma_0) \circledast I(x,y)) - (G(x,y,\sigma_1) \circledast I(x,y)) \tag{6.9}$$

where $I(x,y)$ and $G(x,y,\sigma)$ are defined as in Equation 6.7. Note that the standard deviations of the Gaussian Kernels are swapped, as the roughness is inversely correlated to the image intensity features. The fine Gaussian Kernel with $\sigma_0 = 0.1$ is chosen to extinguish sensor noise. The coarse Gaussian Kernel with $\sigma_1 = 0.5$ is chosen in accordance to fine Gaussian Kernel in 6.7. This ensures that image features on all scales equally affect the augmentation.

Subsequently, I normalized the height variation model to the $[0,1]$ interval and used it as input for the Blender `bump_to_normal` shader node. The maximum strength of the bump map was set to $0.003m$, a seemingly realistic variation of urban ground surfaces, which led to visually pleasing results.

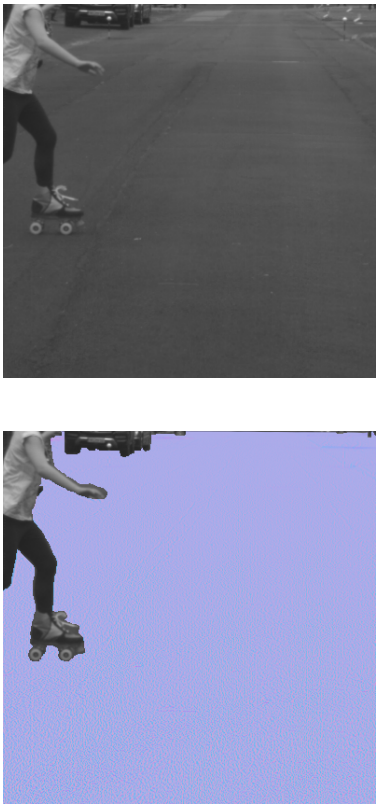

Figure 6.7.: **Top:** Original image for reference. **Bottom:** Same image where the appearance of the ground surface has been replaced with the estimated shading normal. Light blue $(0.5, 0.5, 1.0)$ depicts an upright normal, red colors are tilted towards the x-direction, and green colors are tilted towards the y-direction. Like the roughness parameter, the height variations are derived from difference of Gaussians filtering of the input image, but include higher frequencies. Subsequently, they applied to the ground surface with a $bump_to_normal$ function. Note, that the intensity of the bump map has been amplified for demonstrative purposes.

6.4.3. Separation of Diffuse and Specular Appearance

To augment the ground plane pixels with physically based wetness effects, a separation of the diffuse and specular image intensities needs to be performed. As described in Section 2.3.2, the observed pixel intensity is the result of many light transport paths and can be formulated by the Rendering Equation in 2.4. Under the assumption that the ground plane is non-emissive, the Rendering Equation of the ground plane pixels can be split into two parts: First, all light transport paths that are diffusely reflected by the ground plane pixel. Second, all light transport paths that are reflected specularly by the ground plane pixel.

The specular pixel intensity can be estimated by tracing all specular light transport paths with the estimated specular BRDF parameters of the ground plane:

Figure 6.8.: **Top:** Original image for reference. **Center:** Specular appearance of the image, which is approximated by path tracing the image data in screen-space with the previously estimated BRDF parameters. **Bottom:** Same image brightness enhanced. The specular appearance mostly contains the local intensity maxima, due to the increased roughness, *e.g.*, the bright spot of the sewer lid. However, specular intensity also increases with the increasing visibility of the bright skylight, *e.g.*, note the dark spot in front of the rollerskate that originates from the occluded skylight.

$$\hat{L}_s(X, \omega_o) = \int_\Omega \hat{f}_s(X, \omega_i, \omega_o) L(X, \omega_i) \cos\Theta_i d\omega_i \tag{6.10}$$

where $\hat{L}_s(X, \omega_o)$ is the estimated specular pixel intensity, $\hat{f}_s(X, \omega_i, \omega_o)$ is the estimated specular BRDF model, and all other parameters are chosen as in Equation 2.4.

Subsequently, the diffuse pixel intensity can be estimated by subtracting the estimated specular pixel intensities from the original observed pixel intensities:

$$\hat{L}_d(X, \omega_o) = L(X, \omega_o) - \hat{L}_s(X, \omega_o) \tag{6.11}$$

where $\hat{L}_d(X, \omega_o)$ is the estimated diffuse pixel intensity, $L(X, \omega_o)$ is the original pixel intensity, and $\hat{L}_s(X, \omega_o)$ is the previously estimated specular pixel intensity as in Equation 6.10.

It is apparent that the quality of both estimates depends on the quality of the specular BRDF model and the quality of the specular light transport simulation. The approximation of the specular BRDF parameters is described in Section 6.4.2. In the following, I detail my proposed method to approximate the specular light transport paths in screen-space.

Specular Screen-Space Approximation

Ray tracing and path tracing simulate global lighting, *e.g.*, specular light paths, by following camera light transport path in world-space, *c.f.*, Section 2.3.2. To this end, lighting and reflectance information needs to be available for all scene objects. However, for the proposed augmentation method, lighting and reflectance information is only available within the camera frustrum.

Nevertheless, specular light paths can be approximated in screen-space, a technique that was previously applied for improving the depth estimation of stereo algorithms [107]. In the computer graphics domain, this approximation is known as Screen-Space-Reflections and commonly used to approximation reflections for real-time computer graphics [99]. Here, the input image is projected onto the depth map, and the reflections are ray traced within the camera frustrum.

For the approximation of the specular light transport path, I propose to extend the Screen-Space-Reflections technique, by utilizing the projected input image as in Section 6.4.1 with

the structured and consistent scene representation from Section 6.3. Furthermore, I propose to replace the sparse ray tracing with a dense path tracing, implemented with the Blender Cycles Engine. In contrast to a conventional Screen-Space-Reflection technique, the proposed method better simulates surfaces and objects that are occluded or outside of the camera frustrum. Additionally, the path tracing allows the simulation of multiple light bounces and complex light transport paths.

Figure 6.9.: **Top:** Original image for reference. **Bottom:** The diffuse appearance estimate is obtained by subtracting the specular appearance estimate (*c.f.*, Figure 6.8) from the original image. Note how the overall intensity is decreased compared to the original image and high-frequency details are reduced, just like for real-world surfaces.

6.4.4. Augmentation of the Ground Plane Wetness

For the wetness augmentation of the ground plane I utilize the water-level distribution from Section 6.3.3 and apply it to an adapted variant of the physically based wet surface model by Lagard [76], *c.f.*, Equation 2.5.2.

The water-level distribution $W(u, v, w)$ contains values in the range of $[0, 1]$, where 0.5 corresponds to full water saturation and 1.0 to full water accumulation on the surface. To augment the ground surface with BRDF models for both saturation and accumulation, $W(u, v, w)$ is mapped to the corresponding parameter ranges. More precisely, $W_{sat}(u, v, w) = clamp(2 * W(u, v, w))$ and $W_{sat}(u, v, w) = clamp(2 * W(u, v, w) - 1)$, where $clamp()$ is the clamp function for the range $[0, 1]$. Subsequently, I refer to w_{sat} and w_{acc} as saturation and accumulation for a single pixel location and an arbitrary global wetness parameter.

Ground Plane Saturation Appearance The wet surface model by Lagarde [76] assumes an IOR of $n = 1.5$ as it targets real-time applications, where this approximation is often applied with reasonable quality. However, this results in unrealistically intense reflections for very wet surfaces and unrealistically faint reflections for almost dry surfaces. In addition to modifying the surface roughness, I propose to modify the IOR values also, as the limitations of the real-time engines do not apply to the Blender Cycles Engine. Adapting Equation 2.18, the proposed model for the water saturated ground pixel intensity L_{sat} reads:

$$L_{sat} = lerp(\hat{L}_d, \rho * \hat{L}_d, w_{sat}) + \hat{L}_{s,sat}(lerp(\theta_0, \rho * \theta_0, 0.5 * w_{sat})) \tag{6.12}$$

where $lerp()$ is the linear interpolation function with the third parameter as the blending weight, ρ is the porosity attenuation

factor as in Equation 2.17, $\hat{L}_d$ is the diffuse pixel intensity estimate as in Equation 6.11, θ_0 are the approximated ground pixel roughness and IOR, and $\hat{L}_{s,sat}$ is the newly path traced specular pixel intensity approximation as in Equation 6.10 for the modified BRDF parameters.

Ground Plane Accumulation Appearance Lagarde [76] further proposes to use linear interpolation of the roughness and surface normals to simulate water accumulation, *c.f.*, Section 2.5.2. Following the water saturation case, I propose to extend this interpolation to the IOR parameter and the proposed model for the ground pixel accumulated intensity L_{acc} then reads:

$$L_{acc} = \hat{L}_{d,sat} + \hat{L}_{s,sat}(lerp(\theta_{sat}, \theta_{water}, 0.95 * w_{acc})) \tag{6.13}$$

where $lerp()$ is the linear interpolation function with the third parameter as the blending weight, $L_{d,sat}$ is the diffuse term of the water saturated pixel intensity as in Equation 6.12, θ_{sat} are the water saturated BRDF parameters which this time include the shading normal, θ_{water} are the corresponding parameters of smooth water, and $L_{s,acc}$ is the newly path traced specular pixel intensity approximation as in Equation 6.10 for the modified BRDF parameters. The empirical scaling factor of 0.95 for the blending weight was chosen to limit the specular boost and smoothness of the puddles.

6.4.5. Augmentation of the Water Particles

I modeled the optical scattering properties of the water particle distributions from Section 6.3.4 and 6.3.5, *i.e.*, the droplets on the windshield and water spray and mists, with the principled BSDF shader by Disney [15]. To this end, I used the provided shader implementation by Blender. I assigned the efficient Principled BSDF surface shader to the water droplets, as volume scattering effects within the droplets can be mostly ignored. For the water spray and mists, I used the Principled BSDF volume shader, which is computationally more expensive but can model volume scattering and absorption accurately.

During rendering with the Blender Cycles Engine, the water particle distributions are path traced from the camera, simulating scattering, absorption, and reflection effects accurately. For

the rendering of the water droplets, I adjusted the thin-lens camera model with an aperture of $0.003m$ and a focus distance of $100m$. With this camera setup, I could simulate depth-of-field blurring of the water droplets in 3D-space. However, this effect also applies to the remaining more in-focus scene elements, introducing minimal additional blurring.

6.5. Experiments and Results

The virtual scene setup, described in Section 6.3, is modeled as a Blender [11] scene file. However, the cardboard meshes are newly constructed for each camera pose and take around $4s$ to compute on an Intel Core i7 3770k CPU.

The augmentation method, described in Section 6.4, is implemented in python and can access the scene file via the Blender python interface for parametrizing and rendering a camera pose within the virtual scene. The initial projection of the UV-maps is implemented on the CPU and takes a total of ≈ 1 minute to compute for all scene objects. The subsequent rendering is done with the GPU Cycles Engine with 512 samples per pixel to minimize render noise. Each image without volume effects, *i.e.*, only augmenting ground surface wetness, is computed in $\approx 40 - 50s$ on an NVidia GeForce 1070 GTX graphics card. Augmenting all effects simultaneously increases the render times to $\approx 7minutes$ per frame, due to the computational requirements of the volume scatter effects. With these computation times, the presented method is not suited for on-the-fly augmentation of the training data. Instead, it has to be precomputed and stored on the disc.

The method was applied to three sequences of the HD1K Benchmarking suite. However, I was not able to test the method for network training, as the finalized data and its ground truth were published 01.02.2018, after studies. For an impression about the augmentation of each precipitation effect, I refer to the Appendix A. Subsequently, I analyze the perceptional shortcomings and limitations that I could identify.

6.5.1. Limitations from Dynamic Geometry

The ground surface wetness augmentation creates reflections for the water saturated ground material and more vividly the puddles. The static environment, *i.e.*, buildings, vegetation and background, are all reflected with high perceptional quality, *c.f.*,

(a) Inconsistent Lighting (b) Geometry Artifacts

Appendix A.

However, for the reflections and lighting of the dynamic objects, artifacts can be observed, *c.f.*, Figure 6.10. This can be explained. The point clouds of the underlying datasets only include the static world. Consequently, the geometry of the dynamic objects had to be approximated. A cardboard approach was used, where 2D planes are fitted through the object masks, *c.f.*, Section 6.3.2. While, in most cases, the reflections of the dynamic objects are plausible, the cardboard reflections depend on the accuracy of the labeled masks. When the accuracy of the labeled masks is low, reflection artifacts can occur due to unreasonable cardboard positions. The lighting and shadowing are also affected by the cardboard approximation. Areas below large objects, *e.g.*, cars, are occluded from the skylight and thus shaded. However, due to the cardboard approximation, the light may still be reflected in those areas, creating an unrealistic brightening of the shadows.

6.6. Summary and Outlook

6.6.1. Summary

In this chapter, I presented an augmentation method of precipitation effects for automotive computer vision datasets. The method includes a plausible simulation of ground surface water penetration, reflections from puddles and drenched regions, as well as light scattering from water spray, mists, and water droplets on the windshield. I commenced by describing the process of creating a structured and consistent virtual scene representation from given masks and 3D point clouds of an auto-

motive dataset. I then showed how this scene can be enriched with physically plausible distributions for water accumulation and water particles in 3D space. Finally, I proposed to modify the original image data, utilizing screen-space ray path tracing, a physically based wetness model of ground surfaces, image-based estimates of the material parameters, and the previously created distribution models of water accumulation and particles.

The method was applied to the HD1K dataset due to the availability of high-fidelity point clouds but could be adapted for other datasets as well. The images, produced with the proposed method, show that each precipitation effect is simulated with high perceptional realism, *c.f.*, Appendix A.

6.6.2. Outlook

In the future, an application to neural network training of automotive computer vision tasks is of particular interest. The method has the potential to increase network robustness towards real-world precipitation effects, which are challenging edge cases. Further, the method could be beneficial to increase robustness towards reflections and low contrast situations in general.

I identified two limitations caused by the cardboard approximation of the geometry: First, artifacts in the reflections when the dynamic object masks are inaccurate and the cardboards placed in unreasonable positions. Second, inconsistent lighting in otherwise occluded areas, *e.g.*, below cars. The limitation could be averted by applying a model-based fit to the dynamic object masks or by utilizing datasets with point cloud information of the dynamic objects.

7
Conclusion

Following the presentation of my work in the previous chapters, I have yet to conclude this thesis with a review of my work and an outlook on what I believe remains to be adressed in the context of synthetic reference data generation.

7.1. Summary

One observation was the starting point of my presented work: When using state-of-the-art synthetic datasets to test, train, and validate computer vision algorithms there are systematic differences in comparison to using real-world datasets. The discrepancies can be explained by a low realism of the simulation and inaccuracies of several image formation aspects, *e.g.*, the light transport simulation, the material models, and the texture parameterizations of the materials. Materials are the link between the light sources and sensors and the core element for achieving the goal of synthesizing images that computer vision algorithms cannot distinguish from real-world data.

In Chapter 4, I presented my commencing work of exploring the upper limits of computer graphics and material acquisition [47]. I showed that the realism of the simulation can be sufficiently increased, by combining unbiased light transport algorithms with high-fidelity BTF reflectance data from state-of-the-art measurement devices. The proposed method can, in general, synthesize images that are virtually identical to real-world datasets, both perceptional and in terms of optical flow performance. Only the most reflective materials with high-frequency geometric details require reflectance data, that is captured with a higher sampling density.

However, BTF reflectance data have limitations in terms of rendering performance and suitability for massive datasets. In Chapter 5, I presented my follow-up work on overcoming these limitations [47]. I showed that efficient BRDF models can be fitted to the BTF reflectance data of various material types with

sufficient quality. Perceptional differences to the BTF reference only occur for semi-translucent materials with subsurface scattering and macroscopic shadowing effects. However, these effects pose no challenge to optical flow algorithms and, consequently, the BRDF models match the performance of the BTF reference. On the other hand, reflective material, *e.g.*, metals, and materials with high-frequency surface normals, *e.g.*, coated leather and fabrics, are much more challenging for optical flow. But even those can be replicated with closely matching or even improved accuracy, *i.e.*, for highly reflective materials where the BTF accuracy is limited. In all cases, a physically based Cook-Torrance model, fitted with my proposed technique for surface normal estimation, is best suited for rendering synthetic datasets efficiently.

Not for all tasks and domains, a purely synthetic approach to reference data generation is feasible and cost-effective. One example is the training of neural networks for the automotive domain. Here, the scale and variety of the encountered scenes is massive and can hardly be replicated in entirety by a virtual setup. In Chapter 6, I presented an alternative solution. By applying prior knowledge about physical and optical material properties, I proposed an augmentation method of real-world automotive datasets with precipitation effects. The method simulates wet ground surfaces, water droplets on the windshield, and particles from water spray and mists, by tracing light paths in 3D space. On a perceptional level, the synthesizes images recreate the simulated effects accurately. This opens up the opportunity to utilize the method for increasing network robustness towards the simulated effects as well as reflections and low contrast situations in general.

7.2. Outlook

One major questions regarding reflectance modeling for synthetic datasets is yet to be answered. Many application domains of computer vision algorithms are of massive scale, where it is not feasible to measure the reflectance of the entire scene to create synthetic datasets. Nevertheless, even partial measurements for each material class could increase the confidence of the datasets. To this end, several methods can be used to synthesize larger textures from measurements. One method is to stitch

measured patches to a large tileable texture [29, 30]. However, this approach is only applicable to purely high-frequency materials without macroscopic reflectance features. More sophisticated methods analyze the entire frequency spectrum of the measured textures and synthesize new textures with similar frequency distributions [52]. A more recent approach is to synthesize variations of the measured textures by applying them to baseline textures via style transfer from neural networks [37, 36]. In a future work, it would be interesting to study all three approaches in conjunction with measured reflectance data. A study of resulting periodic patterns and visible seams would be of particular interest, as optical flow and other vision algorithms are sensitive to such artifacts. While style transfer approach seemingly produces excellent results, it additionally needs to be ensured that the resulting textures sufficiently resemble measured data.

Optical Flow was used throughout this thesis as a judge of reflectance data quality. It was chosen, because I deemed it most sensitive to modeling inaccuracies due to its dependency on the statistical distributions of image intensities. However, it would be interesting to see my result applied to other computer vision tasks, *e.g.*, semantic segmentation or object detection. These methods rely more on macroscopic image features and, therefore, potentially have different simulation requirements.

Unfortunately, the HD1K dataset ground truth data was not finalized within the time frame of my studies. Consequently, my presented augmentation method remains to be evaluated in the context of neural network training. Several computer vision tasks are potential candidates, with optical flow and stereo estimation being the most natural candidates, due to the availablity of ground truth data in the underlying dataset. If positively evaluated, the method can complement existing data augmentation techniques for improving the robustness of the neural networks.

A

Appendix

(a) Original Image

(b) $w = 1/3$

(c) $w = 2/3$

(d) $w = 1$

Figure A.1.: Images exemplary augmentations of ground surface wetness as in Chapter 6.
a) Original image without ground surface wetness augmentation image for reference.
(b) Augmented image with a low global wetness level of $w = 1/3$. Ground surface is partially saturated with water and puddles start to form.
(c) Augmented image with a medium global wetness level of $w = 2/3$. Ground surface is almost fully saturated with water and small puddles have accumulated.
(d) Augmented image with maximum global wetness level $w = 1.0$. Ground surface is fully saturated with water and large puddles have accumulated.

(a) Original Image

(b) $w = 1/3$

(c) $w = 2/3$

(d) $w = 1$

Figure A.2.: Images exemplary augmentations of ground surface wetness as in Chapter 6.
a) Original image without ground surface wetness augmentation image for reference.
(b) Augmented image with a low global wetness level of $w = 1/3$. Ground surface is partially saturated with water and puddles start to form.
(c) Augmented image with a medium global wetness level of $w = 2/3$. Ground surface is almost fully saturated with water and small puddles have accumulated.
(d) Augmented image with maximum global wetness level $w = 1.0$. Ground surface is fully saturated with water and large puddles have accumulated.

(a) Original Image

(b) $w = 1/3$

(c) $w = 2/3$

(d) $w = 1$

Figure A.3.: Images exemplary augmentations of ground surface wetness as in Chapter 6.
a) Original image without ground surface wetness augmentation image for reference.
(b) Augmented image with a low global wetness level of $w = 1/3$. Ground surface is partially saturated with water and puddles start to form.
(c) Augmented image with a medium global wetness level of $w = 2/3$. Ground surface is almost fully saturated with water and small puddles have accumulated.
(d) Augmented image with maximum global wetness level $w = 1.0$. Ground surface is fully saturated with water and large puddles have accumulated.

(b) $m = 0$

(b) $m = 1/3$

(c) $m = 2/3$

(d) $m = 1$

Figure A.4.: Images depict exemplary augmentations of water spray and mists from Chapter 6.
a) An image augmented with a maximum global wetness level $w = 1.0$, but no water spray and mists are added.
(b) The same image with a low amount of water spray and mists particles added $m = 1/3$. The average view distance at camera height level is around $450m$.
(c) The same image with a medium amount of water spray and mists particles added $m = 2/3$. The average view distance at camera height level is around $300m$.
(d) The same image with a high amount of water spray and mists particles added $m = 1.0$. The average view distance at camera height level is around $150m$.

Figure A.5.: Images depict exemplary augmentations of water droplets on the windshield from Chapter 6. In all cases the maximum ground surface wetting was also applied.

List of Own Publications

Güssefeld, B., Kondermann, D., Schwartz, C., and Klein, R. "Are reflectance field renderings appropriate for optical flow evaluation?" In: *IEEE International Conference on Image Processing (ICIP)*. IEEE. Paris, France, 2014

Güssefeld, B., Honauer, K., and Kondermann, D. "Creating Feasible Reflectance Data for Synthetic Optical Flow Datasets". In: *Advances in Visual Computing*. Ed. by Bebis, G., Boyle, R., Parvin, B., Koracin, D., Porikli, F., Skaff, S., Entezari, A., Min, J., Iwai, D., Sadagic, A., Scheidegger, C., and Isenberg, T. Cham: Springer International Publishing, 2016, pp. 77–90. ISBN: 978-3-319-50835-1

Kondermann, D., Nair, R., Meister, S., Mischler, W., Güssefeld, B., Honauer, K., Sabine, H., Brenner, C., and Jähne, B. "Stereo Ground Truth With Error Bars". In: *Proc. ACCV*. 2014. URL: `http://hci.iwr.uni-heidelberg.de/Benchmarks/document/StereoErrorBars`

Kondermann, D., Nair, R., Honauer, K., Krispin, K., Andrulis, J., Brock, A., Güssefeld, B., Rahimimoghaddam, M., Hofmann, S., Brenner, C., and Jähne, B. "The HCI Benchmark Suite: Stereo and Flow Ground Truth with Uncertainties for Urban Autonomous Driving". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2016, pp. 19–28. DOI: `10.1109/CVPRW.2016.10`

List of Tables

List of Figures

Bibliography

[1] Abu Alhaija, H., Mustikovela, S. K., Mescheder, L., Geiger, A., and Rother, C. "Augmented Reality Meets Computer Vision: Efficient Data Generation for Urban Driving Scenes". In: *International Journal of Computer Vision* 126.9 (2018), pp. 961–972. ISSN: 1573-1405. DOI: 10.1007/s11263-018-1070-x. URL: https://doi.org/10.1007/s11263-018-1070-x.

[2] Aittala, M., Weyrich, T., and Lehtinen, J. "Two-shot SVBRDF Capture for Stationary Materials". In: *ACM Trans. Graph.* 34.4 (July 2015), 110:1–110:13. ISSN: 0730-0301. DOI: 10.1145/2766967. URL: http://doi.acm.org/10.1145/2766967.

[3] Alhaija, H. A., Mustikovela, S. K., Geiger, A., and Rother, C. "Geometric Image Synthesis". In: *CoRR* abs/1809.04696 (2018). arXiv: 1809.04696. URL: http://arxiv.org/abs/1809.04696.

[4] Aodha, O. M., Humayun, A., Pollefeys, M., and Brostow, G. J. "Learning a Confidence Measure for Optical Flow". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.5 (2013), pp. 1107–1120. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2012.171.

[5] Ashikhmin, M. and Shirley, P. "An Anisotropic Phong BRDF Model". In: *Journal of Graphics Tools* 5.2 (2000), pp. 25–32. DOI: 10.1080/10867651.2000.10487522.

[6] Baker, S., Scharstein, D., Lewis, J. P., Roth, S., Black, M. J., and Szeliski, R. "A Database and Evaluation Methodology for Optical Flow". In: *International Journal of Computer Vision* 92.1 (2011), pp. 1–31. ISSN: 1573-1405. DOI: 10.1007/s11263-010-0390-2. URL: https://doi.org/10.1007/s11263-010-0390-2.

[7] Barron, J. L., Fleet, D. J., and Beauchemin, S. S. "Performance of optical flow techniques". In: *International Journal of Computer Vision* 12.1 (1994), pp. 43–77. ISSN: 1573-1405. DOI: 10.1007/BF01420984. URL: https://doi.org/10.1007/BF01420984.

[8] Becker, F., Lenzen, F., Kappes, J. H., and Schnörr, C. "Variational recursive joint estimation of dense scene structure and camera motion from monocular high speed traffic sequences". In: *International journal of computer vision* 105.3 (2013), pp. 269–297.

[9] Beckmann, P. and Spizzichino, A. *The scattering of electromagnetic waves from rough surface*. Pergamon, 1963.

[10] Belcour, L. and Barla, P. "A Practical Extension to Microfacet Theory for the Modeling of Varying Iridescence". In: *ACM Transactions on Graphics* 36.4 (2017), p. 65. DOI: 10.1145/3072959.3073620. URL: https://hal.archives-ouvertes.fr/hal-01518344.

[11] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Institute, Amsterdam: Blender Foundation, 2014. URL: http://www.blender.org.

[12] Blinn, J. F. "A Generalization of Algebraic Surface Drawing". In: *SIGGRAPH Comput. Graph.* 16.3 (July 1982), pp. 273–. ISSN: 0097-8930. DOI: 10.1145/965145.801290. URL: http://doi.acm.org/10.1145/965145.801290.

[13] Blinn, J. F. “Models of Light Reflection for Computer Synthesized Pictures”. In: *SIGGRAPH Comput. Graph.* 11.2 (1977), pp. 192–198. ISSN: 0097-8930. DOI: 10.1145/965141.563893. URL: http://doi.acm.org/10.1145/965141.563893.

[14] Bosch, A., Zisserman, A., and Munoz, X. “Image Classification using Random Forests and Ferns”. In: *2007 IEEE 11th International Conference on Computer Vision*. 2007, pp. 1–8. DOI: 10.1109/ICCV.2007.4409066.

[15] Burley, B. and Studios, W. D. A. “Physically-based shading at Disney”. In: *Physically Based Shading in Filmand Game Production, ACM SIGGRAPH 2012 Courses*. Vol. 2012. 2012, pp. 1–7. DOI: 10.1145/2343483.2343493. URL: http://selfshadow.com/publications/s2012-shading-course/.

[16] Butler, D. J., Wulff, J., Stanley, G. B., and Black, M. J. “A Naturalistic Open Source Movie for Optical Flow Evaluation”. In: *Computer Vision, ECCV 2012*. Ed. by Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., and Schmid, C. Vol. 7577. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2012, pp. 611–625. ISBN: 978-3-642-33782-6. DOI: 10.1007/978-3-642-33783-3_44.

[17] Buxton, B. and Buxton, H. “Computation of optic flow from the motion of edge features in image sequences”. In: *Image and Vision Computing* 2.2 (1984), pp. 59 –75. ISSN: 0262-8856. DOI: https://doi.org/10.1016/0262-8856(84)90001-5. URL: http://www.sciencedirect.com/science/article/pii/0262885684900015.

[18] Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., and Ranzuglia, G. “MeshLab: an Open-Source Mesh Processing Tool”. In: *Eurographics Italian Chapter Conference*. Ed. by Scarano, V., Chiara, R. D., and Erra, U. The Eurographics Association, 2008. ISBN: 978-3-905673-68-5. DOI: 10.2312/LocalChapterEvents/ItalChap/ItalianChapConf2008/129-136.

[19] Ciresan, D. C., Meier, U., and Schmidhuber, J. “Multi-column Deep Neural Networks for Image Classification”. In: *CoRR* abs/1202.2745 (2012). arXiv: 1202.2745. URL: http://arxiv.org/abs/1202.2745.

[20] Cook, R. L. and Torrance, K. E. “A Reflectance Model for Computer Graphics”. In: *ACM Trans. Graph.* 1.1 (1982), pp. 7–24. ISSN: 0730-0301. DOI: 10.1145/357290.357293.

[21] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., and Schiele, B. “The Cityscapes Dataset for Semantic Urban Scene Understanding”. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 3213–3223. DOI: 10.1109/CVPR.2016.350.

[22] Dana, K. J., Ginneken, B. van, Nayar, S. K., and Koenderink, J. J. “Reflectance and Texture of Real-world Surfaces”. In: *ACM Trans. Graph.* 18.1 (1999), pp. 1–34. ISSN: 0730-0301. DOI: 10.1145/300776.300778. URL: http://doi.acm.org/10.1145/300776.300778.

[23] Debevec, P., Hawkins, T., Tchou, C., Duiker, H.-P., Sarokin, W., and Sagar, M. "Acquiring the Reflectance Field of a Human Face". In: *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '00. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 2000, pp. 145–156. ISBN: 1-58113-208-5. DOI: 10.1145/344779.344855. URL: http://dx.doi.org/10.1145/344779.344855.

[24] Deschaintre, V., Aittala, M., Durand, F., Drettakis, G., and Bousseau, A. "Single-image SVBRDF Capture with a Rendering-aware Deep Network". In: *ACM Trans. Graph.* 37.4 (July 2018), 128:1–128:15. ISSN: 0730-0301. DOI: 10.1145/3197517.3201378. URL: http://doi.acm.org/10.1145/3197517.3201378.

[25] Donath, A. and Kondermann, D. "Is Crowdsourcing for Optical Flow Ground Truth Generation Feasible?" In: Lecture Notes in Computer Science 7963 (2013). Ed. by Chen, M., Leibe, B., and Neumann, B., pp. 193–202. DOI: 10.1007/978-3-642-39402-7_20.

[26] Dosovitskiy, A., Fischery, P., Ilg, E., Häusser, P., Hazirbas, C., Golkov, V., Smagt, P. v. d., Cremers, D., and Brox, T. "FlowNet: Learning Optical Flow with Convolutional Networks". In: *2015 IEEE International Conference on Computer Vision (ICCV)*. 2015, pp. 2758–2766. DOI: 10.1109/ICCV.2015.316.

[27] Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., and Koltun, V. "CARLA: An Open Urban Driving Simulator". In: *Proceedings of the 1st Annual Conference on Robot Learning*. Ed. by Levine, S., Vanhoucke, V., and Goldberg, K. Vol. 78. Proceedings of Machine Learning Research. PMLR, 2017, pp. 1–16. URL: http://proceedings.mlr.press/v78/dosovitskiy17a.html.

[28] Ebert, D. S., Musgrave, F. K., Peachey, D., Perlin, K., and Worley, S. *Texturing and Modeling, Third Edition: A Procedural Approach (The Morgan Kaufmann Series in Computer Graphics)*. 3rd ed. Morgan Kaufmann, 2002. ISBN: 1558608486.

[29] Efros, A. A. and Leung, T. K. "Texture synthesis by non-parametric sampling". In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. Vol. 2. 1999, 1033–1038 vol.2. DOI: 10.1109/ICCV.1999.790383.

[30] Efros, A. A. and Freeman, W. T. "Image Quilting for Texture Synthesis and Transfer". In: *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '01. New York, NY, USA: ACM, 2001, pp. 341–346. ISBN: 1-58113-374-X. DOI: 10.1145/383259.383296. URL: http://doi.acm.org/10.1145/383259.383296.

[31] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. "The Pascal Visual Object Classes (VOC) Challenge". In: *International Journal of Computer Vision* 88.2 (2010), pp. 303–338. ISSN: 1573-1405. DOI: 10.1007/s11263-009-0275-4. URL: https://doi.org/10.1007/s11263-009-0275-4.

[32] Felzenszwalb, P. F. and Huttenlocher, D. P. "Efficient Graph-Based Image Segmentation". In: *International Journal of Computer Vision* 59.2 (2004), pp. 167–181. ISSN: 1573-1405. DOI: 10.1023/B:VISI.0000022288.19776.77. URL: https://doi.org/10.1023/B:VISI.0000022288.19776.77.

[33] Foley, J. D., Dam, A. van, Feiner, S. K., and Hughes, J. F. *Computer Graphics: Principles and Practice in C*. 2nd. Pearson, 1996. ISBN: 8131705056.

[34] Foo, S. C. "A gonioreflectometer for measuring the bidirectional reflectance of material for use in illumination computation". PhD thesis. Cornell University, 1997.

[35] Gaidon, A., Wang, Q., Cabon, Y., and Vig, E. "VirtualWorlds as Proxy for Multi-object Tracking Analysis". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 4340–4349. DOI: 10.1109/CVPR.2016.470.

[36] Gatys, L. A., Ecker, A. S., and Bethge, M. "A Neural Algorithm of Artistic Style". In: *CoRR* abs/1508.06576 (2015). arXiv: 1508.06576. URL: http://arxiv.org/abs/1508.06576.

[37] Gatys, L. A., Ecker, A. S., and Bethge, M. "Texture Synthesis Using Convolutional Neural Networks". In: *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*. NIPS'15. Montreal, Canada: MIT Press, 2015, pp. 262–270. URL: http://dl.acm.org/citation.cfm?id=2969239.2969269.

[38] Geiger, A., Lenz, P., and Urtasun, R. "Are we ready for autonomous driving? The KITTI vision benchmark suite". In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. 2012, pp. 3354–3361. DOI: 10.1109/CVPR.2012.6248074.

[39] Geisler-Moroder, D. and Dür, A. "A New Ward BRDF Model with Bounded Albedo". In: *Computer Graphics Forum* 29.4 (2010), pp. 1391–1398. ISSN: 1467-8659. DOI: 10.1111/j.1467-8659.2010.01735.x. URL: http://dx.doi.org/10.1111/j.1467-8659.2010.01735.x.

[40] Georgiev, I., Křivánek, J., Davidovič, T., and Slusallek, P. "Light Transport Simulation with Vertex Connection and Merging". In: *ACM Trans. Graph.* 31.6 (2012), 192:1–192:10. ISSN: 0730-0301. DOI: 10.1145/2366145.2366211. URL: http://doi.acm.org/10.1145/2366145.2366211.

[41] Goldman, D. B., Curless, B., Hertzmann, A., and Seitz, S. M. "Shape and Spatially-Varying BRDFs from Photometric Stereo". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.6 (2010), pp. 1060–1071. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2009.102.

[42] Goodfellow, I., Bengio, Y., and Courville, A. *Deep Learning*. The MIT Press, 2016. ISBN: 0262035618, 9780262035613.

[43] Gottfried, J.-M. and Kondermann, D. "Charon suite software framework". In: *IPOL 2012 Meeting on Image Processing Libraries*. 2012.

[44] Greivenkamp, J. E. *Field guide to geometrical optics*. Fg01. Bellingham: SPIE, 2004. URL: http://cds.cern.ch/record/2222473.

[45] Gu, J., Tu, C.-I., Ramamoorthi, R., Belhumeur, P., Matusik, W., and Nayar, S. "Time-varying Surface Appearance: Acquisition, Modeling and Rendering". In: *ACM Trans. Graph.* 25.3 (2006), pp. 762–771. ISSN: 0730-0301. DOI: 10.1145/1141911.1141952. URL: http://doi.acm.org/10.1145/1141911.1141952.

[46] Güssefeld, B., Honauer, K., and Kondermann, D. "Creating Feasible Reflectance Data for Synthetic Optical Flow Datasets". In: *Advances in Visual Computing*. Ed. by Bebis, G., Boyle, R., Parvin, B., Koracin, D., Porikli, F., Skaff, S., Entezari, A., Min, J., Iwai, D., Sadagic, A., Scheidegger, C., and Isenberg, T. Cham: Springer International Publishing, 2016, pp. 77–90. ISBN: 978-3-319-50835-1.

[47] Güssefeld, B., Kondermann, D., Schwartz, C., and Klein, R. "Are reflectance field renderings appropriate for optical flow evaluation?" In: *IEEE International Conference on Image Processing (ICIP)*. IEEE. Paris, France, 2014.

[48] Hachisuka, T. and Jensen, H. W. "Stochastic Progressive Photon Mapping". In: *ACM SIGGRAPH Asia 2009 Papers*. SIGGRAPH Asia '09. Yokohama, Japan: ACM, 2009, 141:1–141:8. ISBN: 978-1-60558-858-2. DOI: 10.1145/1661412.1618487. URL: http://doi.acm.org/10.1145/1661412.1618487.

[49] Haeusler, R. and Kondermann, D. "Synthesizing Real World Stereo Challenges". In: *Pattern Recognition: 35th German Conference, GCPR 2013, Saarbrücken, Germany, September 3-6, 2013. Proceedings*. Ed. by Weickert, J., Hein, M., and Schiele, B. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 164–173. ISBN: 978-3-642-40602-7. DOI: 10.1007/978-3-642-40602-7_17. URL: https://doi.org/10.1007/978-3-642-40602-7_17.

[50] Haines, E. and Akenine-Möller, T., eds. *Ray Tracing Gems*. http://raytracinggems.com. Apress, 2019.

[51] Haltakov, V., Unger, C., and Ilic, S. "Framework for Generation of Synthetic Ground Truth Data for Driver Assistance Applications". In: *Pattern Recognition*. Ed. by Weickert, J., Hein, M., and Schiele, B. Vol. 8142. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2013, pp. 323–332. ISBN: 978-3-642-40601-0. DOI: 10.1007/978-3-642-40602-7_35.

[52] Han, C., Risser, E., Ramamoorthi, R., and Grinspun, E. "Multiscale Texture Synthesis". In: *ACM Trans. Graph.* 27.3 (Aug. 2008), 51:1–51:8. ISSN: 0730-0301. DOI: 10.1145/1360612.1360650. URL: http://doi.acm.org/10.1145/1360612.1360650.

[53] Han, J. Y. and Perlin, K. "Measuring Bidirectional Texture Reflectance with a Kaleidoscope". In: *ACM Trans. Graph.* 22.3 (2003), pp. 741–748. ISSN: 0730-0301. DOI: 10.1145/882262.882341. URL: http://doi.acm.org/10.1145/882262.882341.

[54] Hart, J. C. "Perlin Noise Pixel Shaders". In: *Proceedings of the ACM SIGGRAPH/EUROGRAPHICS Workshop on Graphics Hardware*. HWWS '01. New York, NY, USA: ACM, 2001, pp. 87–94. ISBN: 1-58113-407-X. DOI: 10.1145/383507.383531. URL: http://doi.acm.org/10.1145/383507.383531.

[55] Heeger, D. J. "Model for the extraction of image flow". In: *J. Opt. Soc. Am. A* 4.8 (1987), pp. 1455–1471. DOI: 10.1364/JOSAA.4.001455. URL: http://josaa.osa.org/abstract.cfm?URI=josaa-4-8-1455.

[56] Hendrycks, D. and Dietterich, T. G. “Benchmarking Neural Network Robustness to Common Corruptions and Perturbations”. In: *CoRR* abs/1807.01697 (2018). arXiv: 1807.01697. URL: http://arxiv.org/abs/1807.01697.

[57] Hess, M., Koepke, P., and Schult, I. “Optical Properties of Aerosols and Clouds: The Software Package OPAC”. In: *Bulletin of the American Meteorological Society* 79.5 (1998), pp. 831–844. ISSN: 00030007, 15200477. URL: http://www.jstor.org/stable/26215068.

[58] Hirschmuller, H. “Stereo Processing by Semiglobal Matching and Mutual Information”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30.2 (2008), pp. 328–341. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2007.1166.

[59] Hnat, K., Porquet, D., Merillou, S., and Ghazanfarpour, D. “Real-time wetting of porous media”. In: *Machine Graphics and Vision* 15.3/4 (2006), p. 401.

[60] Honauer, K. “Performance Metrics and Test Data Generation for Depth Estimation Algorithms”. PhD thesis. 2019. DOI: 10.11588/heidok.00025758.

[61] Honauer, K., Maier-Hein, L., and Kondermann, D. “The HCI Stereo Metrics: Geometry-Aware Performance Analysis of Stereo Algorithms”. In: *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*. ICCV ’15. Washington, DC, USA: IEEE Computer Society, 2015, pp. 2120–2128. ISBN: 978-1-4673-8391-2. DOI: 10.1109/ICCV.2015.245. URL: http://dx.doi.org/10.1109/ICCV.2015.245.

[62] Horn, B. K. and Schunck, B. G. “Determining optical flow”. In: *Artificial Intelligence* 17.1 (1981), pp. 185 –203. ISSN: 0004-3702. DOI: http://dx.doi.org/10.1016/0004-3702(81)90024-2. URL: http://www.sciencedirect.com/science/article/pii/0004370281900242.

[63] Hosek, L. and Wilkie, A. “An Analytic Model for Full Spectral Sky-dome Radiance”. In: *ACM Trans. Graph.* 31.4 (2012), 95:1–95:9. ISSN: 0730-0301. DOI: 10.1145/2185520.2185591. URL: http://doi.acm.org/10.1145/2185520.2185591.

[64] Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., and Brox, T. “Flownet 2.0: Evolution of optical flow estimation with deep networks”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 2462–2470.

[65] Jakob, W. *Mitsuba renderer*. http://www.mitsuba-renderer.org. 2010.

[66] Jensen, H. W. *Realistic Image Synthesis Using Photon Mapping*. Natick, MA, USA: A. K. Peters, Ltd., 2009. ISBN: 1568814623.

[67] Kajiya, J. T. “The Rendering Equation”. In: *SIGGRAPH Comput. Graph.* 20.4 (1986), pp. 143–150. ISSN: 0097-8930. DOI: 10.1145/15886.15902.

[68] Karis, B. and Games, E. “Real shading in Unreal Engine 4”. In: *Physically Based Shading in Theory and Practice, ACM SIGGRAPH 2013 Courses*. Anaheim, California, 2013. ISBN: 978-1-4503-2339-0. DOI: 10.1145/2504435.2504457.

[69] Kazhdan, M. and Hoppe, H. “Screened Poisson Surface Reconstruction”. In: *ACM Trans. Graph.* 32.3 (2013), 29:1–29:13. ISSN: 0730-0301. DOI: 10.1145/2487228.2487237. URL: http://doi.acm.org/10.1145/2487228.2487237.

[70] Koenderink, J. J. and Doorn, A. J. van. "Affine structure from motion". In: *J. Opt. Soc. Am. A* 8.2 (1991), pp. 377–385. DOI: 10.1364/JOSAA.8.000377. URL: http://josaa.osa.org/abstract.cfm?URI=josaa-8-2-377.

[71] Kondermann, D., Nair, R., Honauer, K., Krispin, K., Andrulis, J., Brock, A., Güssefeld, B., Rahimimoghaddam, M., Hofmann, S., Brenner, C., and Jähne, B. "The HCI Benchmark Suite: Stereo and Flow Ground Truth with Uncertainties for Urban Autonomous Driving". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2016, pp. 19–28. DOI: 10.1109/CVPRW.2016.10.

[72] Kondermann, D., Nair, R., Meister, S., Mischler, W., Güssefeld, B., Honauer, K., Sabine, H., Brenner, C., and Jähne, B. "Stereo Ground Truth With Error Bars". In: *Proc. ACCV*. 2014. URL: http://hci.iwr.uni-heidelberg.de/Benchmarks/document/StereoErrorBars.

[73] Křivánek, J., Georgiev, I., Hachisuka, T., Vévoda, P., Šik, M., Nowrouzezahrai, D., and Jarosz, W. "Unifying Points, Beams, and Paths in Volumetric Light Transport Simulation". In: *ACM Trans. Graph.* 33.4 (2014), 103:1–103:13. ISSN: 0730-0301. DOI: 10.1145/2601097.2601219. URL: http://doi.acm.org/10.1145/2601097.2601219.

[74] Lafortune, E. P. and Willems, Y. D. *Using the modified Phong reflectance model for physically based rendering*. Report CW 197. Celestijnenlaan 200A, 3001 Heverlee, Belgium: Departement Computerwetenschappen, KU Leuven, 1994, p. 19.

[75] Lafortune, E. P. F., Foo, S.-C., Torrance, K. E., and Greenberg, D. P. "Non-linear Approximation of Reflectance Functions". In: *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '97. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1997, pp. 117–126. ISBN: 0-89791-896-7. DOI: 10.1145/258734.258801. URL: https://doi.org/10.1145/258734.258801.

[76] Lagarde, S. *Physicall Based Wet Surfaces*. https://seblagarde.wordpress.com/2013/04/14/water-drop-3b-physically-based-wet-surfaces/. Accessed: 2019-05-17. 2013.

[77] Lagarde, S. and Rousiers, C. de. "Moving Frostbite to Physically Based Rendering". In: *Physically Based Shading in Theory and Practice, ACM SIGGRAPH 2014 Courses*. Vancouver, Canada, 2014. ISBN: 978-1-4503-2962-0. DOI: 10.1145/2614028.2615431. URL: http://doi.acm.org/10.1145/2614028.2615431.

[78] Le Gall, D. "MPEG: A video compression standard for multimedia applications". In: *Communications of the ACM* 34.4 (1991), pp. 46–59.

[79] Lekner, J. and Dorf, M. C. "Why some things are darker when wet". In: *Applied Optics* 27.7 (1988), pp. 1278–1280.

[80] Leyton, M. *A Generative Theory of Shape*. Berlin, Heidelberg: Springer-Verlag, 2001. ISBN: 3-540-42717-1.

[81] Li, T.-M., Aittala, M., Durand, F., and Lehtinen, J. "Differentiable Monte Carlo Ray Tracing Through Edge Sampling". In: *ACM Trans. Graph.* 37.6 (Dec. 2018), 222:1–222:11. ISSN: 0730-0301. DOI: 10.1145/3272127.3275109. URL: http://doi.acm.org/10.1145/3272127.3275109.

[82] Li, Z., Sunkavalli, K., and Chandraker, M. "Materials for Masses: SVBRDF Acquisition with a Single Mobile Phone Image". In: *CoRR* abs/1804.05790 (2018). arXiv: 1804.05790. URL: http://arxiv.org/abs/1804.05790.

[83] Lombardot, B, Luengo-Oroz, M, Melani, C, Faure, E, Santos, A, Peyrieras, N, Ledesma-Carbayo, M, Bourgine, P, Neurobiologie Alfred Fessard, G.-s. de, and Yvette, F. "Evaluation of four 3d non rigid registration methods applied to early zebrafish development sequences". In: *MIAAB MICCAI* (2008).

[84] Lu, D. and Weng, Q. "A survey of image classification methods and techniques for improving classification performance". In: *International Journal of Remote Sensing* 28.5 (2007), pp. 823–870. DOI: 10.1080/01431160600746456. eprint: https://doi.org/10.1080/01431160600746456. URL: https://doi.org/10.1080/01431160600746456.

[85] Lu, J., Georghiades, A. S., Rushmeier, H., Dorsey, J., and Xu, C. "Synthesis of material drying history: Phenomenon modeling, transferring and rendering". In: *ACM SIGGRAPH 2006 Courses*. ACM. 2006, p. 6.

[86] Lucas, B. D. and Kanade, T. "An Iterative Image Registration Technique with an Application to Stereo Vision". In: *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*. IJCAI'81. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1981, pp. 674–679. URL: http://dl.acm.org/citation.cfm?id=1623264.1623280.

[87] Mac Aodha, O., Brostow, G. J., and Pollefeys, M. "Segmenting video into classes of algorithm-suitability". In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2010, pp. 1054–1061. DOI: 10.1109/CVPR.2010.5540099.

[88] Marr, D., Poggio, T., and Brenner, S. "A computational theory of human stereo vision". In: *Proceedings of the Royal Society of London. Series B. Biological Sciences* 204.1156 (1979), pp. 301–328. DOI: 10.1098/rspb.1979.0029.

[89] Marschner, S. R., Westin, S. H., Lafortune, E. P. F., and Torrance, K. E. "Image-based bidirectional reflectance distribution function measurement". In: *Appl. Opt.* 39.16 (2000), pp. 2592–2600. DOI: 10.1364/AO.39.002592. URL: http://ao.osa.org/abstract.cfm?URI=ao-39-16-2592.

[90] Marschner, S. R., Westin, S. H., Lafortune, E. P. F., Torrance, K. E., and Greenberg, D. P. "Image-Based BRDF Measurement Including Human Skin". In: *Rendering Techniques' 99*. Ed. by Lischinski, D. and Larson, G. W. Vienna: Springer Vienna, 1999, pp. 131–144. ISBN: 978-3-7091-6809-7.

[91] Marschner, S. R. "Inverse Rendering for Computer Graphics". AAI9839924. PhD thesis. Ithaca, NY, USA, 1998. ISBN: 0-591-93739-5.

[92] Martull, S., Peris, M., and Fukui, K. "Realistic CG Stereo Image Dataset with Ground Truth Disparity Maps". In: *Technical report of IEICE. PRMU* 111.430 (2012), pp. 117–118. URL: http://ci.nii.ac.jp/naid/110009482347/en/.

[93] Matusik, W. "A Data-driven Reflectance Model". AAI0805702. PhD thesis. Cambridge, MA, USA, 2003.

[94] Mayer, N., Ilg, E., Häusser, P., Fischer, P., Cremers, D., Dosovitskiy, A., and Brox, T. "A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 4040–4048. DOI: 10.1109/CVPR.2016.438.

[95] Mayer, N., Ilg, E., Fischer, P., Hazirbas, C., Cremers, D., Dosovitskiy, A., and Brox, T. "What Makes Good Synthetic Training Data for Learning Disparity and Optical Flow Estimation?" In: *CoRR* abs/1801.06397 (2018). arXiv: 1801.06397. URL: http://arxiv.org/abs/1801.06397.

[96] McAllister, D. K., Lastra, A., and Heidrich, W. "Efficient Rendering of Spatial Bi-directional Reflectance Distribution Functions". In: *Proceedings of the ACM SIGGRAPH/EUROGRAPHICS Conference on Graphics Hardware*. HWWS '02. Saarbrucken, Germany: Eurographics Association, 2002, pp. 79–88. ISBN: 1-58113-580-7. URL: http://dl.acm.org/citation.cfm?id=569046.569057.

[97] McAuley, S., Hill, S., Martinez, A., Villemin, R., Pettineo, M., Lazarov, D., Neubelt, D., Karis, B., Hery, C., Hoffman, N., and Zap Andersson, H. "Physically Based Shading in Theory and Practice". In: *ACM SIGGRAPH 2013 Courses*. SIGGRAPH '13. Anaheim, California: ACM, 2013, 22:1–22:8. ISBN: 978-1-4503-2339-0. DOI: 10.1145/2504435.2504457. URL: http://doi.acm.org/10.1145/2504435.2504457.

[98] McCane, B, Novins, K, Crannitch, D, and Galvin, B. "On Benchmarking Optical Flow". In: *Computer Vision and Image Understanding* 84.1 (2001), pp. 126 –143. ISSN: 1077-3142. DOI: http://dx.doi.org/10.1006/cviu.2001.0930. URL: http://www.sciencedirect.com/science/article/pii/S1077314201909300.

[99] McGuire, M. and Mara, M. "Efficient GPU Screen-Space Ray Tracing". In: *Journal of Computer Graphics Techniques (JCGT)* 3.4 (2014), pp. 73–85. ISSN: 2331-7418. URL: http://jcgt.org/published/0003/04/04/.

[100] Meinhardt-Llopis, E., Pérez, J. S., and Kondermann, D. "Horn-schunck optical flow with a multi-scale strategy". In: *Image Processing On Line* (2013), pp. 151–172. DOI: 10.5201/ipol.2013.20.

[101] Meister, S. and Kondermann, D. "Real versus realistically rendered scenes for optical flow evaluation". In: *Electronic Media Technology (CEMT), 2011 14th ITG Conference on*. 2011, pp. 1–6. URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5936557\&tag=1.

[102] Müller, T., Gross, M., and Novák, J. "Practical Path Guiding for Efficient Light-Transport Simulation". In: *Computer Graphics Forum* 36.4 (2017), pp. 91–100. DOI: 10.1111/cgf.13227.

[103] Movshovitz-Attias, Y., Kanade, T., and Sheikh, Y. "How Useful Is Photo-Realistic Rendering for Visual Learning?" In: *Computer Vision – ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III*. Ed. by Hua, G. and Jégou, H. Cham: Springer International Publishing, 2016, pp. 202–217. ISBN: 978-3-319-

49409-8. DOI: 10.1007/978-3-319-49409-8_18. URL: https://doi.org/10.1007/978-3-319-49409-8_18.

[104] Mérillou, S., Dischler, J.-M., and Ghazanfarpour, D. "A BRDF Postprocess to Integrate Porosity on Rendered Surfaces". In: *IEEE Trans. Vis. Comput. Graph.* 6.4 (2000), pp. 306–318. DOI: http://www.computer.org/tvcg/tg2000/v0306abs.htm.

[105] Müller, G. "Data-Driven Methods for Compression and Editing of Spatially Varying Appearance". Dissertation. 2009.

[106] Müller, G., Meseth, J., Sattler, M., Sarlette, R., and Klein, R. "Acquisition, Synthesis, and Rendering of Bidirectional Texture Functions". In: vol. 24. 1. Blackwell Publishing Ltd., 2005, pp. 83–109. DOI: 10.1111/j.1467-8659.2005.00830.x.

[107] Nair, R., Fitzgibbon, A., Kondermann, D., and Rother, C. "Reflection Modeling for Passive Stereo". In: *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*. ICCV '15. Washington, DC, USA: IEEE Computer Society, 2015, pp. 2291–2299. ISBN: 978-1-4673-8391-2. DOI: 10.1109/ICCV.2015.264. URL: http://dx.doi.org/10.1109/ICCV.2015.264.

[108] Nakamae, E., Kaneda, K., Okamoto, T., and Nishita, T. "A Lighting Model Aiming at Drive Simulators". In: *SIGGRAPH Comput. Graph.* 24.4 (1990), pp. 395–404. ISSN: 0097-8930. DOI: 10.1145/97880.97922. URL: http://doi.acm.org/10.1145/97880.97922.

[109] Nakamura, Y., Matsuura, T., Satoh, K., and Ohta, Y. "Occlusion detectable stereo-occlusion patterns in camera matrix". In: *Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 1996, pp. 371–378. DOI: 10.1109/CVPR.1996.517099.

[110] Ngan, A., Durand, F., and Matusik, W. "Experimental Analysis of BRDF Models". In: *Eurographics Symposium on Rendering (2005)*. Ed. by Bala, K. and Dutre, P. The Eurographics Association, 2005. ISBN: 3-905673-23-1. DOI: 10.2312/EGWR/EGSR05/117-126.

[111] Ångström, A. "The Albedo of Various Surfaces of Ground". In: *Geografiska Annaler* 7 (1925), pp. 323–342. ISSN: 16513215. URL: http://www.jstor.org/stable/519495.

[112] Nicodemus, F. E. "Directional Reflectance and Emissivity of an Opaque Surface". In: *Appl. Opt.* 4.7 (1965), pp. 767–775. DOI: 10.1364/AO.4.000767. URL: http://ao.osa.org/abstract.cfm?URI=ao-4-7-767.

[113] Onkarappa, N. and Sappa, A. D. "Synthetic sequences and ground-truth flow field generation for algorithm validation". In: *Multimedia Tools and Applications* (2013), pp. 1–15. ISSN: 1380-7501. DOI: 10.1007/s11042-013-1771-7.

[114] Otte, M. and Nagel, H. H. "Optical flow estimation: Advances and comparisons". In: *Computer Vision — ECCV '94: Third European Conference on Computer Vision Stockholm, Sweden, May 2–6, 1994 Proceedings, Volume I*. Ed. by Eklundh, J.-O. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994, pp. 49–60. ISBN: 978-3-540-48398-4. DOI: 10.1007/3-540-57956-7_5. URL: https://doi.org/10.1007/3-540-57956-7_5.

[115] Pal, N. R. and Pal, S. K. "A review on image segmentation techniques". In: *Pattern Recognition* 26.9 (1993), pp. 1277 –1294. ISSN: 0031-3203. DOI: https://doi.org/10.1016/0031-3203(93)90135-J. URL: http://www.sciencedirect.com/science/article/pii/003132039390135J.

[116] Papenberg, N., Bruhn, A., Brox, T., Didas, S., and Weickert, J. "Highly Accurate Optic Flow Computation with Theoretically Justified Warping". In: *International Journal of Computer Vision* 67.2 (2006), pp. 141–158. ISSN: 1573-1405. DOI: 10.1007/s11263-005-3960-y. URL: https://doi.org/10.1007/s11263-005-3960-y.

[117] Patow, G. and Pueyo, X. "A Survey of Inverse Rendering Problems". In: *Computer Graphics Forum* 22.4 (2003), pp. 663–687. DOI: 10.1111/j.1467-8659.2003.00716.x.

[118] Pharr, M., Jakob, W., and Humphreys, G. "Physically Based Rendering: From Theory to Implementation (3rd ed.)" In: (2016), p. 1200. URL: http://infoscience.epfl.ch/record/220021.

[119] Phong, B. T. "Illumination for Computer Generated Pictures". In: *Commun. ACM* 18.6 (1975), pp. 311–317. ISSN: 0001-0782. DOI: 10.1145/360825.360839. URL: http://doi.acm.org/10.1145/360825.360839.

[120] Pishchulin, L., Jain, A., Wojek, C., Andriluka, M., Thormählen, T., and Schiele, B. "Learning people detection models from few training samples". In: *CVPR 2011*. 2011, pp. 1473–1480. DOI: 10.1109/CVPR.2011.5995574.

[121] Pope, R. M. and Fry, E. S. "Absorption spectrum (380–700 nm) of pure water. II. Integrating cavity measurements". In: *Applied optics* 36.33 (1997), pp. 8710–8723.

[122] Prados, E. and Faugeras, O. "Shape From Shading". In: *Handbook of Mathematical Models in Computer Vision*. Ed. by Paragios, N., Chen, Y., and Faugeras, O. Boston, MA: Springer US, 2006, pp. 375–388. ISBN: 978-0-387-28831-4. DOI: 10.1007/0-387-28831-7_23. URL: https://doi.org/10.1007/0-387-28831-7_23.

[123] Prakash, A., Boochoon, S., Brophy, M., Acuna, D., Cameracci, E., State, G., Shapira, O., and Birchfield, S. "Structured Domain Randomization: Bridging the Reality Gap by Context-Aware Synthetic Data". In: *CoRR* abs/1810.10093 (2018). arXiv: 1810.10093. URL: http://arxiv.org/abs/1810.10093.

[124] Ramanan, D., Felzenszwalb, P., and McAllester, D. "A discriminatively trained, multiscale, deformable part model". In: *2008 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*. Vol. 00. 2008, pp. 1–8. DOI: 10.1109/CVPR.2008.4587597. URL: doi.ieeecomputersociety.org/10.1109/CVPR.2008.4587597.

[125] Ren, S., He, K., Girshick, R., and Sun, J. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". In: *Advances in Neural Information Processing Systems 28*. Ed. by Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., and Garnett, R. Curran Associates, Inc., 2015, pp. 91–99. URL: http://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf.

[126] Ros, G., Sellart, L., Materzynska, J., Vazquez, D., and Lopez, A. M. "The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 3234–3243. DOI: 10.1109/CVPR.2016.352.

[127] Rozenberg, G. and Salomaa, A. *Mathematical Theory of L Systems*. Orlando, FL, USA: Academic Press, Inc., 1980. ISBN: 0125971400.

[128] Ruiters, R., Schwartz, C., and Klein, R. "Example-based Interpolation and Synthesis of Bidirectional Texture Functions". In: *Computer Graphics Forum* 32 (2013), pp. 361 –370. DOI: 10.1111/cgf.12056.

[129] Schlick, C. "An Inexpensive BRDF Model for Physically-based Rendering". In: *Computer Graphics Forum* 13.3 (1994), pp. 233–246. ISSN: 1467-8659. DOI: 10.1111/1467-8659.1330233. URL: http://dx.doi.org/10.1111/1467-8659.1330233.

[130] Schwartz, C., Weinmann, M., Ruiters, R., and Klein, R. "Integrated High-quality Acquisition of Geometry and Appearance for Cultural Heritage". In: *Proceedings of the 12th International Conference on Virtual Reality, Archaeology and Cultural Heritage*. VAST'11. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2011, pp. 25–32. ISBN: 978-3-905674-34-7. DOI: 10.2312/VAST/VAST11/025-032. URL: http://dx.doi.org/10.2312/VAST/VAST11/025-032.

[131] Schwartz, C., Sarlette, R., Weinmann, M., and Klein, R. "DOME II: A Parallelized BTF Acquisition System". In: *Proceedings of the Eurographics 2013 Workshop on Material Appearance Modeling: Issues and Acquisition*. MAM '13. Zaragoza, Spain: Eurographics Association, 2013, pp. 25–31. ISBN: 978-3-905674-48-4. DOI: 10.2312/MAM.MAM2013.025-031.

[132] 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". In: *Sensors (Basel, Switzerland)* 14.5 (2014), 7753—7819. ISSN: 1424-8220. DOI: 10.3390/s140507753. URL: http://europepmc.org/articles/PMC4063053.

[133] Sellent, A, Kondermann, D., Simon, S., Baker, S., Dedeoglu, G., Erdler, O., Parsonage, P., Unger, C., and Niehsen, W. "Optical Flow Estimation versus Motion Estimation". In: (2012).

[134] Shi, J. and Malik, J. "Normalized Cuts and Image Segmentation". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 22.8 (2000), pp. 888–905. ISSN: 0162-8828. DOI: 10.1109/34.868688. URL: https://doi.org/10.1109/34.868688.

[135] Shirley, P. and Wang, C. "Direct Lighting Calculation by Monte Carlo Integration". In: *Photorealistic Rendering in Computer Graphics*. Ed. by Brunet, P. and Jansen, F. W. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994, pp. 52–59. ISBN: 978-3-642-57963-9.

[136] Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M., and Moore, R. "Real-time Human Pose Recognition in Parts from Single Depth Images". In: *Commun. ACM* 56.1 (2013), pp. 116–124. ISSN: 0001-0782. DOI: 10.1145/2398356.2398381. URL: http://doi.acm.org/10.1145/2398356.2398381.

[137] Smith, B. "Geometrical shadowing of a random rough surface". In: *IEEE Transactions on Antennas and Propagation* 15.5 (1967), pp. 668–671. ISSN: 0018-926X. DOI: 10.1109/TAP.1967.1138991.

[138] Standards, U. S. N. B. of and Nicodemus, F. E. *Geometrical considerations and nomenclature for reflectance*. Vol. 160. US Department of Commerce, National Bureau of Standards, 1977.

[139] Sun, B., Sunkavalli, K., Ramamoorthi, R., Belhumeur, P., and Nayar, S. "Time-varying BRDFs". In: *Proceedings of the Second Eurographics Conference on Natural Phenomena*. NPH'06. Vienna, Austria: Eurographics Association, 2006, pp. 15–23. ISBN: 3-905673-38-X. DOI: 10.2312/NPH/NPH06/015-024. URL: http://dx.doi.org/10.2312/NPH/NPH06/015-024.

[140] Tang, P., Huber, D., Akinci, B., Lipman, R., and Lytle, A. "Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques". In: *Automation in Construction* 19.7 (2010), pp. 829 –843. ISSN: 0926-5805. DOI: https://doi.org/10.1016/j.autcon.2010.06.007. URL: http://www.sciencedirect.com/science/article/pii/S0926580510000907.

[141] Tang, Y. and Eliasmith, C. "Deep Networks for Robust Visual Recognition". In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*. ICML'10. Haifa, Israel: Omnipress, 2010, pp. 1055–1062. ISBN: 978-1-60558-907-7. URL: http://dl.acm.org/citation.cfm?id=3104322.3104456.

[142] Taylor, J., Shotton, J., Sharp, T., and Fitzgibbon, A. "The vitruvian manifold: Inferring dense correspondences for one-shot human pose estimation". In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2012, pp. 103–110.

[143] Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., To, T., Cameracci, E., Boochoon, S., and Birchfield, S. "Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization". In: *CoRR* abs/1804.06516 (2018). arXiv: 1804.06516. URL: http://arxiv.org/abs/1804.06516.

[144] Tsirikoglou, A., Kronander, J., Wrenninge, M., and Unger, J. "Procedural Modeling and Physically Based Rendering for Synthetic Data Generation in Automotive Applications". In: *CoRR* abs/1710.06270 (2017). arXiv: 1710.06270. URL: http://arxiv.org/abs/1710.06270.

[145] Twomey, S. A., Bohren, C. F., and Mergenthaler, J. L. "Reflectance and albedo differences between wet and dry surfaces". In: *Appl. Opt.* 25.3 (1986), pp. 431–437. DOI: 10.1364/AO.25.000431. URL: http://ao.osa.org/abstract.cfm?URI=ao-25-3-431.

[146] Vasiljevic, I., Chakrabarti, A., and Shakhnarovich, G. "Examining the Impact of Blur on Recognition by Convolutional Networks". In: *arXiv e-prints* (Nov. 2016). arXiv: 1611.05760 [cs.CV].

[147] Vaudrey, T., Rabe, C., Klette, R., and Milburn, J. "Differences between stereo and motion behaviour on synthetic and real-world stereo sequences". In: *Image and Vision Computing New Zealand, 2008. IVCNZ 2008. 23rd International Conference*. 2008, pp. 1–6. DOI: 10.1109/IVCNZ.2008.4762133.

[148] VEACH, E. "Metropolis light transport". In: *Proc. ACM SIGGRAPH '97* (1997), pp. 65–76. URL: https://ci.nii.ac.jp/naid/10025522170/en/.

[149] Veach, E. and Guibas, L. "Bidirectional Estimators for Light Transport". In: *Photorealistic Rendering Techniques*. Ed. by Sakas, G., Müller, S., and Shirley, P. Berlin, Heidelberg: Springer Berlin Heidelberg, 1995, pp. 145–167. ISBN: 978-3-642-87825-1.

[150] Viola, P. and Jones, M. "Rapid Object Detection using a Boosted Cascade of Simple Features". In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001(CVPR)*. Vol. 01. 2001, p. 511. DOI: 10.1109/CVPR.2001.990517. URL: doi.ieeecomputersociety.org/10.1109/CVPR.2001.990517.

[151] Vorba, J., Karlík, O., Šik, M., Ritschel, T., and Křivánek, J. "On-line Learning of Parametric Mixture Models for Light Transport Simulation". In: *ACM Trans. Graph.* 33.4 (2014), 101:1–101:11. ISSN: 0730-0301. DOI: 10.1145/2601097.2601203. URL: http://doi.acm.org/10.1145/2601097.2601203.

[152] Walter, B., Marschner, S. R., Li, H., and Torrance, K. E. "Microfacet models for refraction through rough surfaces". In: *Proceedings of the 18th Eurographics conference on Rendering Techniques*. Eurographics Association. 2007, pp. 195–206.

[153] Ward, G. J. "Measuring and Modeling Anisotropic Reflection". In: *SIGGRAPH Comput. Graph.* 26.2 (1992), pp. 265–272. ISSN: 0097-8930. DOI: 10.1145/142920.134078.

[154] Weidlich, A. and Wilkie, A. "Arbitrarily layered micro-facet surfaces". In: *Proceedings of the 5th international conference on Computer graphics and interactive techniques in Australia and Southeast Asia*. ACM. 2007, pp. 171–178.

[155] Weidlich, A. and Wilkie, A. "Exploring the Potential of Layered BRDF Models". In: *ACM SIGGRAPH ASIA 2009 Courses*. SIGGRAPH ASIA '09. Yokohama, Japan: ACM, 2009, 7:1–7:58. DOI: 10.1145/1665817.1665824. URL: http://doi.acm.org/10.1145/1665817.1665824.

[156] Weinmann, M., Gall, J., and Klein, R. "Material Classification Based on Training Data Synthesized Using a BTF Database". In: *Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part III*. Ed. by Fleet, D., Pajdla, T., Schiele, B., and Tuytelaars, T. Cham: Springer International Publishing, 2014, pp. 156–171. ISBN: 978-3-319-10578-9. DOI: 10.1007/978-3-319-10578-9_11. URL: https://doi.org/10.1007/978-3-319-10578-9_11.

[157] Weinzaepfel, P., Revaud, J., Harchaoui, Z., and Schmid, C. "DeepFlow: Large displacement optical flow with deep matching". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2013, pp. 1385–1392.

[158] Werlberger, M., Pock, T., Unger, M., and Bischof, H. "Optical flow guided TV-L 1 video interpolation and restoration". In: *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*. Springer. 2011, pp. 273–286.

[159] Wilkie, A., Tobler, R. F., and Purgathofer, W. "Combined Rendering of Polarization and Fluorescence Effects". In: *Rendering Techniques 2001*. Ed. by Gortler, S. J. and Myszkowski, K. Vienna: Springer Vienna, 2001, pp. 197–204. ISBN: 978-3-7091-6242-2.

[160] 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*. Berlin, Heidelberg: Springer-Verlag, 1997, pp. 13–22. ISBN: 3-211-83001-4. URL: http://dl.acm.org/citation.cfm?id=647651.731971.

[161] Wrenninge, M. and Unger, J. "Synscapes: A Photorealistic Synthetic Dataset for Street Scene Parsing". In: *CoRR* abs/1810.08705 (2018). arXiv: 1810.08705. URL: http://arxiv.org/abs/1810.08705.

[162] Wu, C. "Towards Linear-Time Incremental Structure from Motion". In: *2013 International Conference on 3D Vision - 3DV 2013*. 2013, pp. 127–134. DOI: 10.1109/3DV.2013.25.

[163] Xu, L., Jia, J., and Matsushita, Y. "Motion detail preserving optical flow estimation". In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2010, pp. 1293–1300. DOI: 10.1109/CVPR.2010.5539820.

[164] Yilmaz, A., Javed, O., and Shah, M. "Object Tracking: A Survey". In: *ACM Comput. Surv.* 38.4 (Dec. 2006). ISSN: 0360-0300. DOI: 10.1145/1177352.1177355. URL: http://doi.acm.org/10.1145/1177352.1177355.

[165] Yilmaz, A., Javed, O., and Shah, M. "Object Tracking: A Survey". In: *ACM Comput. Surv.* 38.4 (2006). ISSN: 0360-0300. DOI: 10.1145/1177352.1177355. URL: http://doi.acm.org/10.1145/1177352.1177355.

[166] Zach, C., Pock, T., and Bischof, H. "A duality based approach for realtime TV-L 1 optical flow". In: *Joint pattern recognition symposium*. Springer. 2007, pp. 214–223.

[167] Zhang, H. and Voss, K. J. "Bidirectional reflectance study on dry, wet, and submerged particulate layers: effects of pore liquid refractive index and translucent particle concentrations". In: *Appl. Opt.* 45.34 (2006), pp. 8753–8763. DOI: 10.1364/AO.45.008753. URL: http://ao.osa.org/abstract.cfm?URI=ao-45-34-8753.

[168] Zhang, Y., Qiu, W., Chen, Q., Hu, X., and Yuille, A. L. "UnrealStereo: A Synthetic Dataset for Analyzing Stereo Vision". In: *CoRR* abs/1612.04647 (2016). arXiv: 1612.04647. URL: http://arxiv.org/abs/1612.04647.

[169] Zhou Wang, Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. "Image quality assessment: from error visibility to structural similarity". In: *IEEE Transactions on Image Processing* 13.4 (2004), pp. 600–612. ISSN: 1057-7149. DOI: 10.1109/TIP.2003.819861.