

An Interactive Shader for Natural Diffraction Gratings

Bachelorarbeit

der Philosophisch-naturwissenschaftlichen Fakultät
der Universität Bern

vorgelegt von

Michael Single

2014

Leiter der Arbeit:
Prof. Dr. Matthias Zwicker
Institut für Informatik und angewandte Mathematik

Abstract

Inhaltsverzeichnis

1	Introduction	1
1.1	Motivation	1
1.2	Goals	3
1.3	Previous work	4
1.4	Overview	5
2	Theoretical Background	6
2.1	Basics in Modeling Light in Computer Graphics	6
2.1.1	Radiometry	6
2.1.2	Spectral Energy	6
2.1.3	Spectral Power	7
2.1.4	Spectral Irradiance	7
2.1.5	Spectral Radiance	8
2.1.6	BRDF	9
2.1.7	Wavespectrum and Colors	10
2.1.8	Colorspace	11
2.1.9	Spectral Rendering	13
2.2	Wave Theory for Light and Diffraction	13
2.2.1	Basics in Wave Theory	13
2.2.2	Wave Interference	15
2.2.3	Wave Coherence	16
2.2.4	Huygen's Principle	17
2.2.5	Waves Diffraction	18
2.3	Stam's BRDF formulation	20
3	Derivations	26
3.1	Problem Statement and Challenges	26
3.2	Rendering Equation	27
3.3	Adaption of Stam's BRDF discrete height fields	28
3.4	Reproduce FT by DTFT	29
3.5	Special Coherence and Windowing	30
3.6	Reproduce DTFT by DFT	30
3.7	Relative Reflectance	31
3.8	Optimization using Taylor Series	33
3.9	Spectral Rendering	36
3.10	Alternative Approach	37
3.10.1	PQ factors	37
3.10.2	Interpolation	40

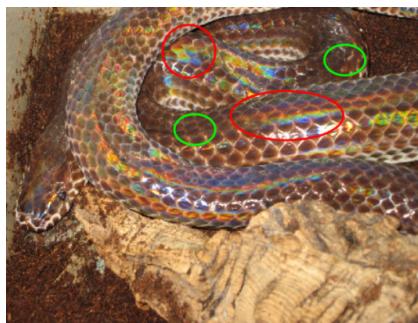
A Signal Processing Basics	42
A.1 Fourier Transformation	42
A.2 Convolution	44
A.3 Taylor Series	44
B Summary of Stam's Derivations	45
C Appendix	48
C.1 Schlick's approximation	48
C.2 Spherical Coordinates	48
C.3 Tangent Space	49
List of Tables	50
List of Figures	50
List of Algorithms	51
Bibliography	52

Kapitel 1

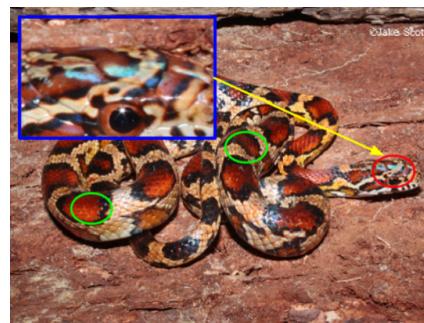
Introduction

1.1 Motivation

As a human being, we visually perceive and experience our whole world in terms of colors, resulting from various kinds of physical light interaction phenomena. Particularly in biology, there are basically two main causes for color production. Firstly, due to pigmentation, which occurs since certain molecules in a biological structure selectively absorb or reflect specific wavelengths from an incident light source. And secondly because of structural colors which are the result of physical interaction of light with a nanostructure, exclusively relying on the structuring of the material and not any other property. One particular example for such biological color production are the colors we can see when having a closer look at the illuminated skin of snake species, as shown in figure 1.1.



(a) Xenopeltis snake



(b) Elaphe Guttata snake

Abbildung 1.1: Examples of pigmentation color (green circles) and structural color (red circles) on snake species¹.

Comparing both snake species with each other, we observe that the Xenopeltis species expresses structural colors in form of iridescent patterns along its scales way stronger than the Elaphe species does. The reason for this lies on the

¹image source of figure 1.1(a) http://www.snakes-alive.co.uk/gallery_5.html and figure 1.1(b) http://www.the-livingrainforest.co.uk/living/view_price.php?id=464

nanostructure of their skins. A natural diffraction grating is a semitransparent layer of biological nano-structures which exhibits a certain degree of regularity to produce structural colors by diffracting an incident light source. There are a vast amount of additional reasons for producing structural colors in nature, such as thin film interference, intra-cellular photonic crystals or diffraction gratings. More detailed examples are listed in figure 1.2.

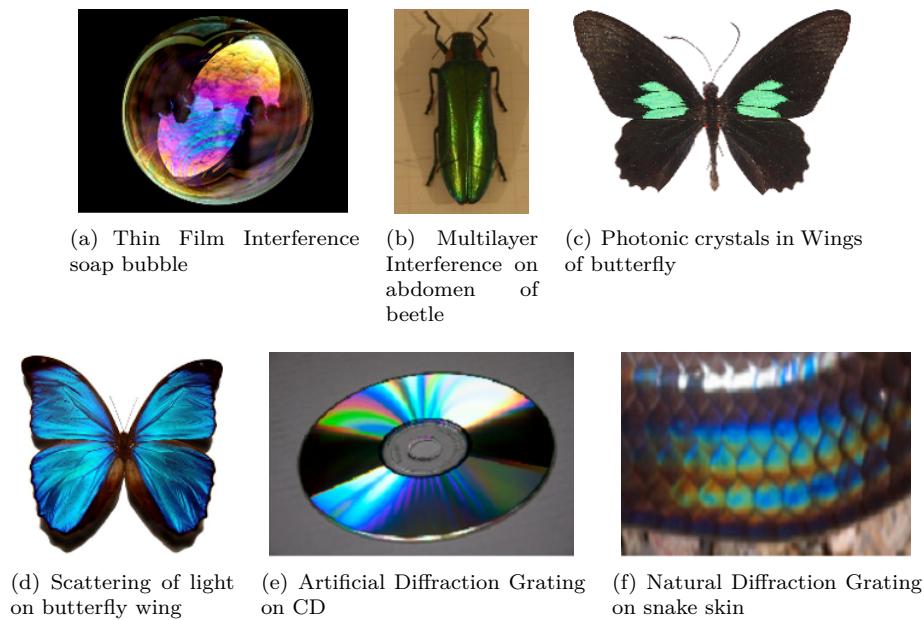


Abbildung 1.2: Examples² for structural colors on the wings and the abdomen of insects, liquids, synthetic structures, and on scales on the skin of reptiles.

Already in the 17th century Robert Hooke was able to relate the cause of structural colors to the microstructure of a material. During his examinations of peacock feathers he found that the colors could be made disappear by wetting the feathers and further observed that the feathers are made of tiny ridges. Building on top of the latest knowledge about interference, Newton related structural colors with wave interference. Nowadays, in the field of computer graphics, many researchers have been attempted rendering structural colors but unfortunately, most currently available approaches are not able to perform interactive rendering or are oversimplified and thus cannot model accurately the effect of diffraction.

²image source of figure:

- 1.2(a): http://www.ualberta.ca/~pogosyan/teaching/PHYS_130/FALL_2010/lectures/lect33/lecture33.html
- 1.2(b): <http://www.itp.uni-hannover.de/~zawischa/ITP/multibeam.html>
- 1.2(c): http://upload.wikimedia.org/wikipedia/commons/a/a4/Parides_sesostris_MHNT_dos.jpg
- 1.2(d): From paper [MT10], figure 6.
- 1.2(e): <http://cnx.org/content/m42496/latest/?collection=col11428/latest>
- 1.2(f): http://www.snakes-alive.co.uk/gallery_5.html

1.2 Goals

The purpose of this thesis is firstly, to simulate physically accurate structural colors caused by the effect of diffraction on various biological structures and secondly implement this simulation as a renderer with interactive behaviour. We mainly focus on structural colors generated by natural diffraction gratings. In particular the approach presented in this thesis applies to surfaces with quasi-periodic structures at the nanometer scale which can be represented as height fields stored in gray-scale images.

Natural gratings like this are found on the scale of reptiles, wings of butterflies or the bodies of various insects but we restrict ourself and focus on snake skins. The height field data of used snake skins are acquired using atomic force microscopy (AFM) and provided by LANE lab at Geneva³.

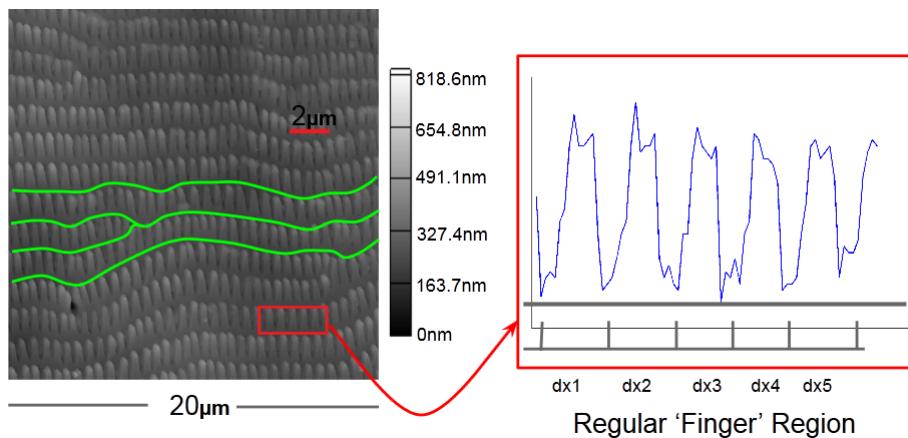


Abbildung 1.3: Height field of a Xenopeltis snake⁴ skin taken by AFM and stored as a grayscale image. Even this natural grating consists locally of regularly aligned (red box) finger-like substructures, we observe globally a curved alignment (green curves).

The renderer discussed in this thesis is based on the pioneering work of J. Stam about diffraction shaders [Sta99] in which he formulated a BRDF modelling the effect of diffraction. Nevertheless we have to adapt his BRDF model since his model assumes, that a given surface of a grating can either be formulated by an analytical function, and has therefore an closed form solution or is that simple that it satisfies relying on statistical methods in order to model the surface. However, in our case, we are dealing with natural diffraction grating, given as an explicit height field as shown in figure 1.3. Hence we neither can rely on image statistics, since the details of natural gratings are to complex nor are we able to formulate an closed form solution. Therefore, this results of this thesis can be considered as an extension of J. Stam's work for the case someone is provided by a explicit height field of a quasi-periodic surface.

³Laboratory of Artificial and Natural Evolution. Website: www.lanevol.org

⁴This image was provided by the LANE lab in Geneva

1.3 Previous work

The whole study of structural colors already started in the the 17th century, Since the first scientific descriptions on structural colors by Hooke in 1665 in his book Micrographia[R.H12]. Hooke investigated feathers of peacocks using one of the first microscopes from this time and found out that the colors on the feather were canceled out whenever a drop of water moistened the feather. He proposed the speculation that a layer of thin plates and air were responsible for reflecting the light and thus he related the structure of the feather to colors. In Newton's book Opticks[I.N14] he described that the colors of the peacock feather are related to the thinness of the transparent part of the feathers. Around 1800 T.Young related Structural colors as a result of wave interference by his two-slit experiment⁵, published in the journal Philosophical Transactions of the Royal Society these times.

In the field of computer graphics J.Stam[Sta99] was the first one who was able to develop reflection models based on wave optics capturing the effect of diffraction. His model is a an approximation of far field diffraction⁶ effects relying on the Kirchhof integral⁷. For a certain class of surfaces which can be modelled as a height field he provides an analytical solution of the BRDF model he proposes. The main idea of his model is the formulate a BRDF as the Fourier Transform applied on the given height field. However, the height field Stam is dealing with are either extremely regular or can be considered as a probabilistic superposition of bumps forming a periodic like structure relying on probabilistic distribution theory⁸. Both height field assumptions allow him to derive an analytical solution using statistical models. However, the height field we are dealing with are measured, complex, biological nano-structures and thus far apart from being globally very regular, as already demonstrated in figure 1.3. It is also not sufficient superimposing one particular finger (considering it as a bump) since it would not capture the complexity of the measured structure. Therefore, we cannot directly rely on his BRDF model when we want to perform interactive rendering since we do not have an analytical solution and hence are not able to solve the Fourier Transform of the given height field on the fly.

In 2012 Cuypers et all [CT12] proposed a wave based Bidirectional scattering distribution function (BSDF)⁹ denoted as WBSDF. Using the rendering equation and Wigner Distribution Functions¹⁰ (WDF) they related their WBSDF model to the incoming wavefront and hence, their model can be adapted such that it can be rendered by a Monte Carlo renderer. The advantage of their model over Stam's is that their models also captures near field diffraction effects. A disadvantage their model is computational expensive since the WDF of a two dimensional surface is a four dimensional function and therefore can hardly be used in order to perform interactive rendering.

⁵See http://en.wikipedia.org/wiki/Double-slit_experiment

⁶See http://en.wikipedia.org/wiki/Fraunhofer_diffraction

⁷See http://en.wikipedia.org/wiki/Kirchhoff_integral_theorem

⁸See http://en.wikipedia.org/wiki/Probability_distribution

⁹See http://en.wikipedia.org/wiki/Bidirectional_scattering_distribution_function

¹⁰See http://en.wikipedia.org/wiki/Wigner_distribution_function

The work of Lindsay and Agu [CT12] an approach in order to perform interactive rendering diffraction effect by precompute and store their BRDF model using spherical harmonics. Nonetheless, for complex natural gratings their BRDF may be insufficient accurate since their approach is using low order spherical harmonics.

1.4 Overview

The remainder of this thesis is organised as the follows: Due to the fact that this thesis has a rather advanced mathematical complexity, the first part of chapter 2 introduces some important definitions about modelling light in computer graphics and some wave theory. These concepts are required in order to be able to follow our later derivations. The second part of chapter 2 contains brief summary of J. Stam's Paper about diffraction shaders, since his BRDF formulation is the basis of our derivations.

In chapter 3 we adapt Stam's BRDF model step-wise in a way that we will end up with a representation which can be implemented in an interactive diffraction renderer when using natural diffraction gratings. We also propose an alternative formulation, the so called PQ approach.

Chapter 4 addresses the practical part of this thesis, the implementation of our diffraction model, explaining all precomputation steps and how rendering is performed in our developed framework for this thesis.

Chapter 5 gives some further insight about diffraction by explaining the topic about diffraction grating in depth. Furthermore, within this chapter we evaluate the qualitative validity of our BRDF models applied on different surface gratings by computing their reflectance and comparing this to the grating equation under similar conditions.

Chapter 6 presents our rendered results, first the so called BRDF maps for all our gratings and shading approaches under various shading parameters and then the actual renderings on a snake mesh. And finally chapter 7 contains the conclusion of this thesis which starts by a review briefly by discussing what has been achieved in this thesis and all the arising drawbacks. Also, there are some words about my personal experience during this thesis.

Kapitel 2

Theoretical Background

2.1 Basics in Modeling Light in Computer Graphics

2.1.1 Radiometry

One purpose of Computer Graphics is to simulate the interaction of light on a surface and how a real-world observer, such as a human eye, will perceive this. These visual sensations of an eye are modeled relying on a virtual camera which captures the emitted light from the surface. The physical basis to measure such reflected light depicts radiometry which is about measuring the electromagnetic radiation transferred from a source to a receiver.

Fundamentally, light is a form of energy propagation, consisting of a large collection of photons, whereat each photon can be considered as a quantum of light that has a position, direction of propagation and a wavelength λ . A photon travels at a certain speed $v = \frac{c}{n}$, that depends only the speed of light c and the refractive index n through which it propagates. Its frequency is defined by $f = \frac{v}{\lambda}$ and its carried amount of energy q , measured in the SI unit Joule, is given by $q = hf = \frac{hv}{\lambda n}$ where h is the Plank's constant. The total energy of a large collection of photons is hence $Q = \sum_i q_i$.

2.1.2 Spectral Energy

It is important to understand that the human eye is not equally sensitive to all wavelength of the spectrum of light and therefore responds differently to specific wavelengths. Remember that our goal is to model the human visual perception. This is why we consider the energy distribution of a light spectrum rather than considering the total energy of a photon collection since then we could weight the distribution according the human visual system. So the question we want to answer is: How is the energy distributed across wavelengths of light?

The idea is to make an energy histogram from a given photon collection. For this we have to order all photons by their associated wavelength, discretize wavelength spectrum, count all photons which then will fall in same wavelength-

interval, and then, finally, normalize each interval by the total energy Q . This will give us a histogram which tells us the spectral energy Q_λ for a given discrete λ interval and thus models the so called spectral energy distribution¹.

2.1.3 Spectral Power

Rendering an image in Computer Graphics corresponds to capturing the color sensation of an illuminated, target scene at a certain point in time. As previously seen, each color is associated by a wavelength and is directly related to a certain amount of energy. In order to determine the color of a to-be-rendered pixel of an image, we have to get a sense of how much light (in terms of energy) passes through the area which the pixel corresponds to. One possibility is to consider the flow of energy $\Phi = \frac{\Delta Q}{\Delta t}$ transferred through this area over a small period of time. This allows us to measure the energy flow through a pixel during a certain amount of time.

In general, power is the estimated rate of energy production for light sources and corresponds to the flux. It is measured in the unit Watts, denoted by Q . Since power is a rate over time, it is well defined even when energy production is varying over time. As with Spectral Energy for rendering, we are really interested in the spectral power $\Phi_\lambda = \frac{\Phi}{\lambda}$, measured in Watts per nanometer.

2.1.4 Spectral Irradiance

Before we can tell how much light is reflected from a given point on a surface towards the viewing direction of an observer, we first have to know how much light arrives at this point. Since in general a point has no length, area or even volume associated, let us instead consider an infinitesimal area ΔA around a such a point. Then, we can ask ourself how much light falls in such a small area. When further observing this process over a short period in time, this quantity is the spectral irradiance E as illustrated in figure 2.1. Summarized, this quantity tells us how much spectral power is incident on a surface per unit area and mathematically is equal:

$$E = \frac{\Phi_\lambda}{\Delta A} \quad (2.1)$$

¹Intensive quantities can be thought of as density functions that tell the density of an extensive quantity at an infinitesimal point.

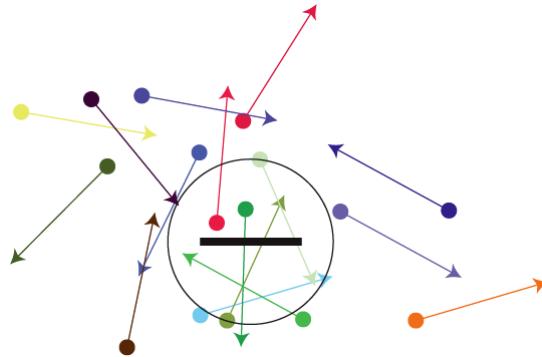
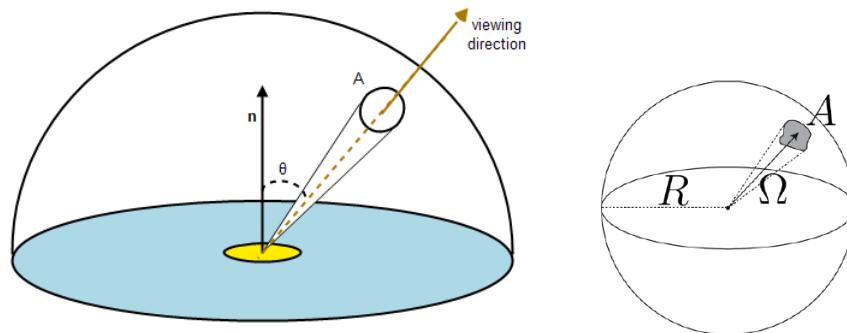


Abbildung 2.1: Irradiance is the summed up radiance over all directions

2.1.5 Spectral Radiance

When rendering an image we have to determine the color of each pixel of the image. Although irradiance tells us how much light is arriving at a point as illustrated in figure 2.1, it tells us little about the direction that light comes from. This relates to how the human eye perceives the brightness of an illuminated objects when looking at it in a certain direction.



(a) Radiances is the density of photons per area per solid angle

(b) Solid angle is the area of a surface patch on a sphere with radius R which is spanned by a set of directions

This concept is described by the radiometric quantity radiance. Basically, this is a measure of light energy passing through or is emitted off from a small area around a point on a surface towards a given direction during a short period in time. More formally this is the spectral power emerging from an arbitrary point (an infinitesimal area around this point) and falls within a given solid angle (see figure² 2.2(b)) in specific direction (usually towards the observer) as shown in figure 2.2(a). Formally, this leads us to the following mathematical formalism:

$$L_\lambda(\omega) = \frac{d^2\Phi_\lambda}{dAd\Omega} \approx \frac{\Phi_\lambda}{\Omega A} \quad (2.2)$$

²Similar figure like used in computer graphics class 2012 in chapter colors

where L is the observed spectral radiance in the unit energy per unit area per solid angle, which is $\text{Wm}^{-2}\text{sr}^{-1}$ in direction ω which has an angle θ between the surface normal and ω , Θ is the total flux or power emitted, θ is the angle between the surface normal and the specified direction, A is the area of the surface and Ω is the solid angle in the unit steradian subtended by the observation or measurement.

It is useful to distinguish between radiance incident at a point on a surface and excitant from that point. Terms for these concepts sometimes used in the graphics literature are surface radiance L_r for the radiance *reflected* from a surface and field radiance L_i for the radiance *incident* at a surface.

2.1.6 BRDF

In order to render the colorization of an observed object, a natural question in computer graphics is what portion of the reflected, incident light a viewer will receive, when he looks at an illuminated object. Therefore for any given surfaces which is illuminated from a certain direction ω_i , we can ask ourself how much light is reflected off of any point on this surface towards a viewing direction ω_r . This is where the Bidirectional Reflectance Distribution Function (short: BRDF) comes into play, which is a radiometric quantity telling us how much light is reflected at an opaque surface. Mathematically speaking, the BRDF is the ratio of the reflected radiance pointing to the direction ω_r to the incident irradiance comming from the inverse direction of ω_i as illustrated in figure 2.2. Hence the BRDF is a four dimensional function defined by four angles θ_i , ϕ_i , θ_r and ϕ_r .

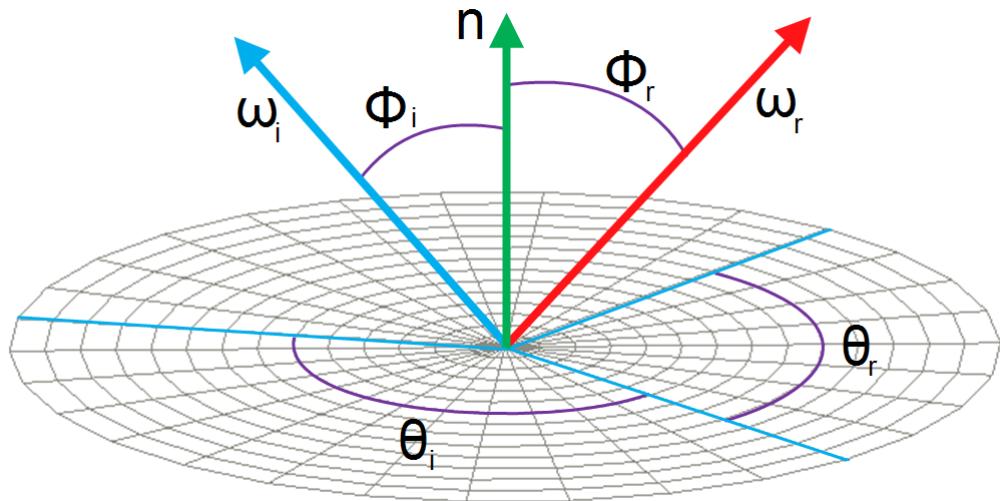


Abbildung 2.2: Illustration of the BRDF model, where ω_i is pointing to the light source and the existing direction is denoted by ω_r . Both direction unit direction vectors defined w.r.t a surface normal \mathbf{n} for every point on the surface.

Which formally is for any given wavelength λ equivalent to:

$$\begin{aligned} BRDF_{\lambda}(\omega_i, \omega_r) &= \frac{dL_r(\omega_r)}{dE_i(\omega_i)} \\ &= \frac{dL_r(\omega_r)}{L_i(\omega_i)\cos(\theta_i)d\omega_i} \end{aligned} \quad (2.3)$$

Where L_r is the reflected spectral radiance, E_i is the spectral irradiance and θ_i is the angle between ω_i and the surface normal \mathbf{n} .

2.1.7 Wavespectrum and Colors

In order to see how crucial the role of human vision plays, let us consider the following definition of color by *Wyszeckiu and Siles*³ stating that *Color is the aspect of visual perception by which an observer may distinguish differences between two structure-free fields of view of the same size and shape such as may be caused by differences in the spectral composition of the radiant energy concerned in the observation.* Therefore, similarly like the humans' perceived sensation of smell and taste, color vision is just another individual sense of perception giving us the ability to distinguish different frequency distribution of light experienced as color.

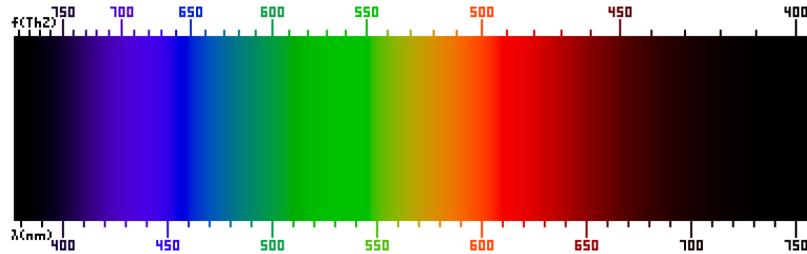


Abbildung 2.3: Frequency (top) and wavelength (bottom) of colors of the visible light spectrum⁴.

In general an eye consists of photoreceptor cells which are responsible for providing ability of color-perception. A schematic of an eye is illustrated in figure 2.4. Basically, there are two specialized types of photoreceptor cells, cone cells which are responsible for color vision and rod cells, which allow an eye to perceive different brightness levels.

³mentioned in Computer Graphics Fundamentals Book from the year 2000

⁴Similar figure like used in computer graphics class 2012 in chapter colors

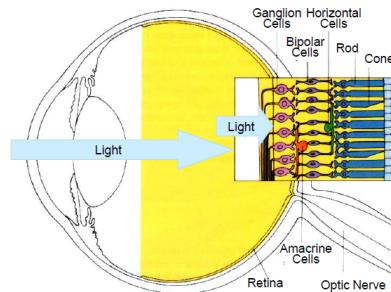


Abbildung 2.4: Schematic⁵ of photoreceptor cells, cones and rods, in human eye

A human eye is made of three different types of cone cells, having their peak sensitivity in sensing color at different wavelength ranges. More precisely, there are cone cells most sensitive to short wavelengths which are between 420nm and 440nm , those which are most sensitive in the middle range between 530nm and 550nm and those which have their peak in the long range, from 560nm to 580nm . In principle, any color sensation in human color perception as shown in figure 2.3 can therefore be described by just three parameters, corresponding to levels of stimulus of the three types of cone cells.

2.1.8 Colorspace

In order to render accurately images of how a human observer sees its world, a mathematical model of the human color perception is required. Remember that color sensation is due to a visual stimulus processed by cone cells in an eye. A human eye contains three different types of cone cells. Therefore, one possible approach is to describe each kind of these cone cells as a function of wavelength, returning a certain intensivity. In the early 1920, from a series of experiments the so called CIE XYZ color space was derived, describing response of cone cells of an average human individual, the so called standard observer. Basically, a statistically sufficiently large number of probands were exposed to different target light colors expressed by their wavelength. The task of each proband was to reproduce these target colors by mixing three given primary colors, red-, green- and blue-light. The strength of each primary color could be manually adjusted by setting their relative intensivity. Those adjustment weights have been measured, aggregated and averaged among all probands for each primary color. This model describes each color as a triple of three real valued numbers⁶, the so called tristimulus values.

Pragmatically speaking, color spaces describes the range of colors a camera can see, a printer can print or a monitor can display. Thus, formally we can

⁵image of illustration has been taken from wikipedia

⁶note that there are negative color weights possible in the CIE XYZ colors space. This is why some human perceived color sensations could not be reconstructed using just an additive color model (adding three positively weighted primary values). Therefore, a probabant was also allowed to move one of the primary colors to the target color and instead was supposed to reproduce this new color mix using the two remaining primaries (subtractive model). The value of the selected, moved primary was then interpreted as being negative weighted in an additive color model.

define it as a mapping a range of physically produced colors from mixed light to an objective description of color sensations registered in the eye of an observer in terms of tristimulus values.

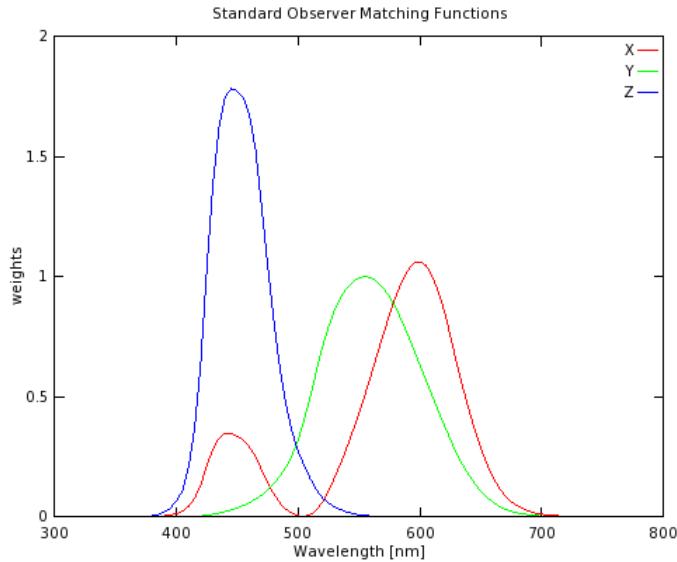


Abbildung 2.5: Plots of our color matching functions we used for rendering

Interpolating all measured tristimuli values gives us three basis functions, the CIE color matching functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, $\bar{z}(\lambda)$. In figure 2.5 are the numerical description of the chromatic response of the observer. They can be thought of as the spectral sensitivity curves of three linear light detectors yielding the CIE Tristimulus values X, Y and Z.

The tristimulus values for a color with a spectral power distribution $I(\lambda)$, are given in terms of the standard observer by:

$$\begin{aligned} X &= \int_{\Lambda} I(\lambda) \bar{x}(\lambda) d\lambda \\ Y &= \int_{\Lambda} I(\lambda) \bar{y}(\lambda) d\lambda \\ Z &= \int_{\Lambda} I(\lambda) \bar{z}(\lambda) d\lambda \end{aligned} \quad (2.4)$$

Where λ , is the wavelength of the equivalent monochromatic light spectrum $\Lambda = [380nm, 780nm]$. Note taht it is not possible to build a display that corresponds to the CIE XYZ colorspace. For this reasons it is necessary to design other color spaces, which are physical realizable, offers efficient encoding, are perceptual uniform and have an intuitive color specification. There are simple conversions between XYZ color space, to other color space described as linear transformations.

2.1.9 Spectral Rendering

When rendering an image, most of the time we are using colors described in a certain RGB color space. However, a RGB colorspace results from a colorspace transformation of the tristimulus values, which themselves are inherent to the human visual system. Therefore, many physically light phenomenon are poorly modeled when always relying on RGB colors for rendering. Using only RGB colors for rendering is alike we would assume that a given light source emits light of only one particular wavelength. But in reality this is barely the case. Spectral rendering is referring to use a certain wavelength spectrum, e.g. the human visible light spectrum, instead simply using the whole range of RGB values in order to render an illuminated scene. This captures the physical reality of specific light sources way more accurate. Keep in mind that, even when we make use of a spectral rendering approach, we have to convert the final spectra to RGB values, when we want to display an image on an actual display.

2.2 Wave Theory for Light and Diffraction

2.2.1 Basics in Wave Theory

In order prepare the reader for physical relevant concepts used during later derivations and reasonings within this thesis, I am going to provide a quick introduction to the fundamental basics of wave theory and related concepts. In physics a wave describes a disturbance that travels from one location to another through a certain medium. The disturbance temporarily displaces the particles in the medium from their rest position which results in an energy transport along the medium during wave propagation. Usually, when talking about waves we are actually referring to a complex valued function which is a solution to the so called wave equation which is modeling how the wave disturbance proceeds in space during time.

There are two types of waves, mechanical waves which deform their mediums during propagation like sound waves and electromagnetic waves consisting of periodic oscillations of an electromagnetic field such as light for example. Like simplified illustrated in figure 2.6, there are several properties someone can use and apply in order to compare and distinguish different waves:

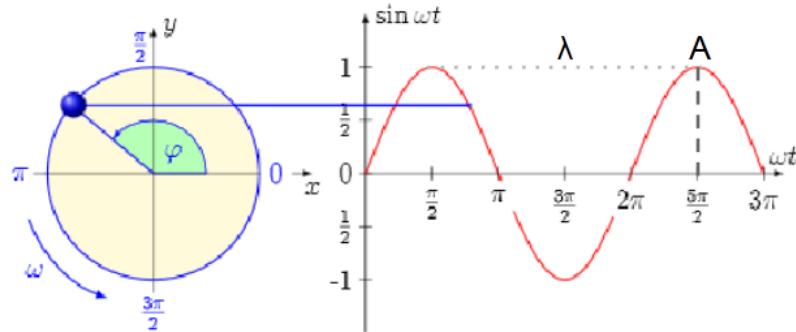


Abbildung 2.6: Simplified, one dimensionaly real valued wave function⁷, giving an idea about some important wave properties. We denote the crest of a wave as the hightest point relative to the equilibrium line (zero height along time axis) and similarly the trough as the lowest point.

Wavelength: Is usually denoted by λ and is a measure for the spatial distance from one point to another until the shape of a wave repeats

Amplitude: Is denoted by A and there are two possible interpretations: First, it is a measure of the height from the equilibrium point to the heighest point of a crest or the lowest point of a trough. This mean the amplitude can be positive or negative. However, usually, someone is just interested the absolute value of an amplitude, the magnitude of a wave. For light waves it is a relative measure of intensity or brightnes to other lugh waves of the same wavelength. And secondly, it can be interpreted as a measure how much energy a wave carries wherate the greater the absoulte aplitude value, the bigger the amount of energy being carried.

Frequency: Is a measure of the number of waves which are passing through a particular point in the propagation medium during a certain time and is denoted by f .

Phase: Is denoted by ϕ . Describes either the offset of initial position of a wave or the relative displacement between or among waves having the same frequency. Two waves two waves with same frequency are denoted by being in phase if they have the same phase. This means they line up everywhere. As a remark, we denote by ω the angular frequency which is equal $2\pi f$.

A geometrical property of waves is their wavefront. This is either a surface or line along the path of wave propagation on which the disturbance at every point has the same phase. Three are basically three types of wavefronts: spherical-, cylindrical- and plane wavefront. If point in a isotropic medium is sending out waves in three dimensions, then the coresponding wavefronts are spheres, centered on the source point. Hence spherical wavefront is the result of a spherical wave, also denoted as a wavelet. Note that for electromagnetic waves, the phase is a poision of a point in time on a wavefront cycle (motion of wave over a whole wavelength) wherat a complete cycle is defined as being equal 360 degrees.

⁷Image source: <http://neutrino.ethz.ch/Vorlesung/FS2013/index.php/vorlesungsskript>

2.2.2 Wave Interference

Next, after having seen that a wave is simply a traveling disturbance along a medium, having some special properties, someone could ask what happens when there are several waves traveling on the same medium. Especially, we are interested how these waves will interact with each other. In physics the term interference denotes the interaction of waves when they encounter each other at a point along their propagation medium. At each point where two waves superpose, their total displacement at these points is the sum of the displacements of each individual wave at those points. Then, the resulting wave is having a greater or lower amplitude than each separate wave and this we can interpret the interference as the addition operator for waves. Two extreme scenarios are illustrated in figure 2.7. There are basically three variants of interferences which can occur, depending on how crest and troughs of the waves are matched up:

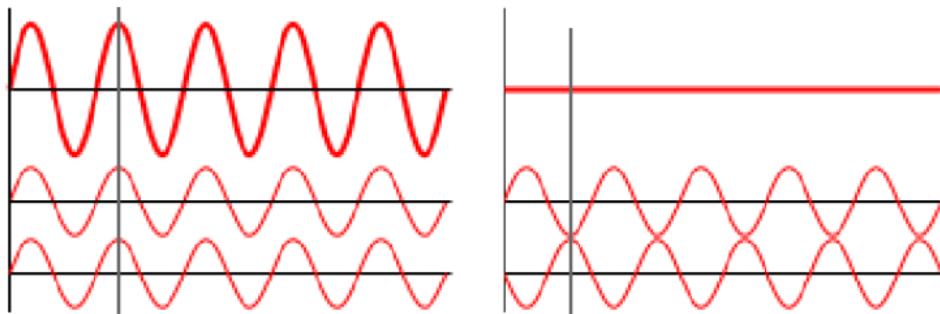


Abbildung 2.7: Interference scenarios⁸ when two waves meet: On the left hand side, there is constructive interference and on the right hand side there is destructive interference illustrated.

- Either a crest of a wave meets a crest of another wave or similarly a trough meets a trough of another wave. This scenario is denoted as constructive interference and occurs at any location along the medium where the two interfering waves have a displacement in the same direction. This is equivalent like saying that the phase difference between the waves is a multiple of 2π . Then the resulting amplitude at that point is being much larger than the amplitude of an individual wave. For two waves with an equal amplitude interfering constructively, the resulting amplitude is twice as large as the amplitude of an individual wave.
- Either a crest of a wave meets a trough of another wave or vice versa. This scenario is denoted as destructive interference and occurs at any location along the medium where the two interfering waves have a displacement in the opposite direction. This is like saying that the phase difference between the waves is an odd multiple of π . Then the waves completely cancel each other out at any point they superimpose.

⁸Image source: [http://en.wikipedia.org/wiki/Interference_\(wave_propagation\)](http://en.wikipedia.org/wiki/Interference_(wave_propagation))

- If the phase difference between two waves is intermediate between the first two scenarios, then the magnitude of the displacement lies between the minimal and maximal values which we could get from constructive interference.

Keep in mind that when two or more waves interfere with each other, the resulting wave will have a different frequency. For a wave, having a different frequency also means having a different wavelength. Therefore, this directly implies that a light of a different color, than its source waves have, is emitted.

2.2.3 Wave Coherence

When considering waves which are traveling on a shared medium along the same direction, we could examine how their phase difference is changing over time. Formulating the change of their relative phase as a function of time will provide us a quantitative measure of the synchronism of two waves, the so called wave coherence. In order to better understand this concept, let us consider a perfectly mathematical sine wave and second wave which is a phase-shifted replica of the first one. A property of mathematical waves is that they keep their shape over an infinity amount of moved wavelengths. In our scenario, both waves are traveling along the same direction on the same medium, like exemplarily illustrated in figure 2.8.

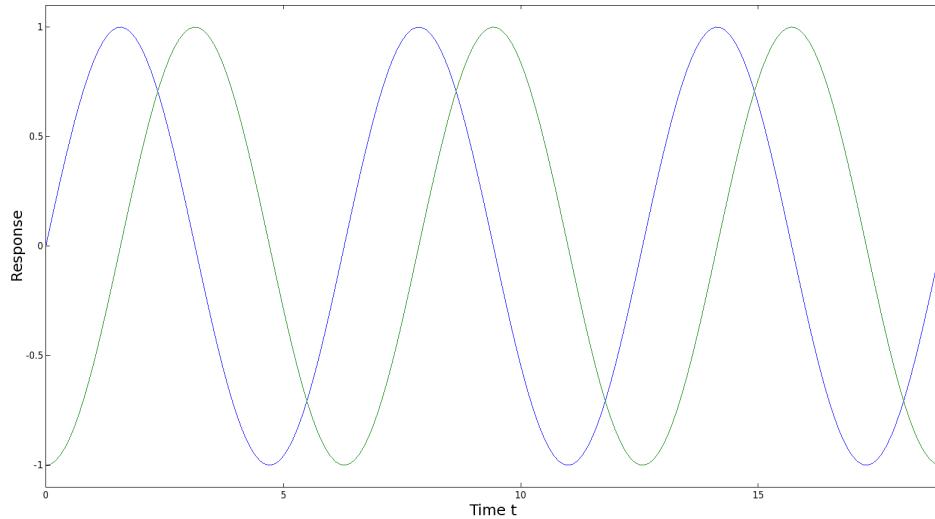


Abbildung 2.8: Two mathematical sine waves which are perfectly coherent which means that their phase difference is constant for every point in time.

Taking the difference between the two sine waves from the previous figure yields always a constant number. Therefore, those two waves are said to be coherent and hence perfectly synchronous over time. Notice that this scenario is completely artificial since in nature there are no mathematical sine waves. Rather, the phase difference is then a function of time $p(t)$. The more coherent two waves are, the slower this function will change over time. In fact, two waves are said

to be coherent if they are either of the same frequency, temporally in phase or have the same amplitude at every point in time. Thus two waves are coherent if they are generated at the same time, having the same frequency, amplitude, and phase. Reversely, Waves are considered incoherent or also asynchronous if they have no stable phase difference. This means $p(t)$ is heavily varying over time. Coherence describes the effect of whether waves will tend to interfere with each other constructively or destructively at a certain point in time and space. Thus this is a property of waves that enables stationary interference. The more correlated two waves are, the higher their degree of coherence is. In physics coherence between waves is quantified by the cross-correlation function, which basically predicts the value of a second wave using the value of the first one. There are two basic coherence classifications:

- Spatial coherence is dealing with the question of what is the range of distance between two points in space in the extend of a wave for which there is occurring a significant effect of interference when averaged over time. This is formally answered by considering the correlation between waves at different point in space. The range of distance is also denoted as the coherence area.
- Temporal coherence examines the ability of how well a wave will interfere with itself at different moments in time. Mathematically, this kind of coherence is computed by averaging the measured correlation between the value of the wave and the delayed version of itself at different pairs of time. The Coherence time denotes the time for which the propagating wave is coherent and we therefore can predict its phase using the correlation function. The distance a wave has traveled during the coherence time is denoted as the coherence length.

2.2.4 Huygen's Principle

Besides from a wave's phase and amplitude, also its propagation directly affects the interaction between different waves and how they could interfere with each other. This is why it makes sense to formulate a model which allows us to predict the position of a moving wavefront and how it moves in space. This is where Huygen's Principle comes into play. It states that any each point of a wavefront may be regarded as a point source that emits spherical wavelets in every direction. Within the same propagation medium, these wavelets travel at the same speed as their source wavefront. The position of the new wavefront results by superimposing all of these emitted wavelets. Geometrically, the surface that is tangential to the secondary waves can be used in order to determine the future position of the wavefront. Therefore, the new wavefront encloses all emitted wavelets. Figure 2.9 visualizes Huygen's Principle for a wavefront reflected off from a plane surface.

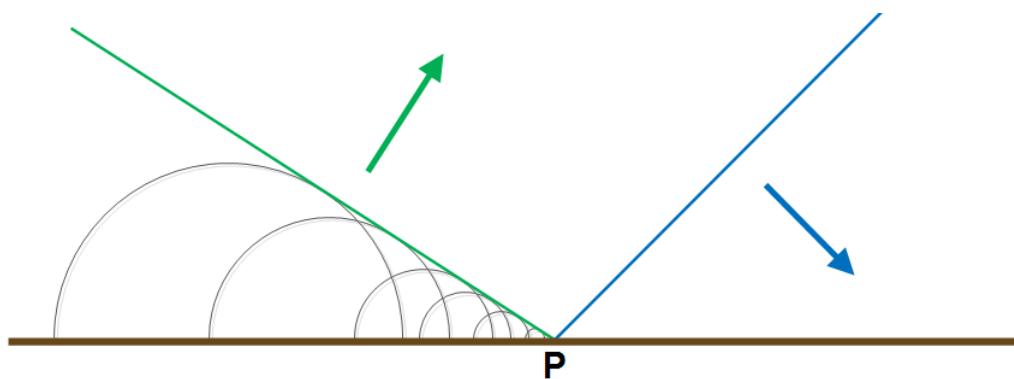


Abbildung 2.9: A moving wavefront (blue) encounters an obstacle (a surface in brown colors) and produces a new wavefront (green) as a result of superposition among all secondary wavelets.

2.2.5 Waves Diffraction

Revisiting Hugen's Principle we know that each point on a wavefront can be considered as a source of a spherical wavelet which propagates in every direction. But what exactly happens when a wave's propagation direction is occluded by an object? What will be the outcome when applying Huygen's Principle for that case? An example scenario for this case is shown in figure 2.10.

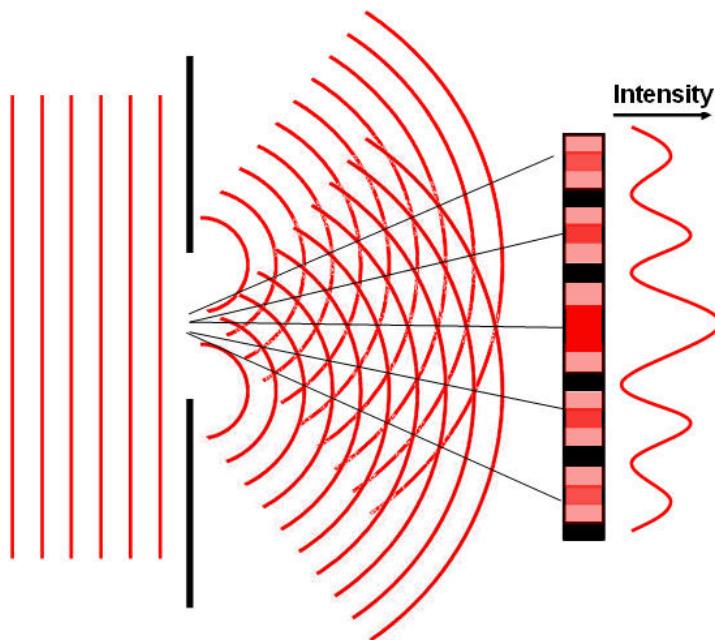


Abbildung 2.10: Illustration⁹ of a diffraction scenario in which a plane wavefront passes through a surface with a certain width and how the wave will be bent, also showing the intensity of the resulting wave.

Whenever a propagating wavefront is partially occluded by an obstacle, the wave is not only moving in the direction along its propagation, but is also bent around the edges of the obstacle. In physics, this phenomenon is called diffraction. Waves are diffracted due to interference which occurs among all wavelets when applying Huygen's Principle for the case when a wavefront hits an obstacle. Generally, the effect of diffraction is most expressed for waves whose wavelength is roughly similar in size to the dimension of the occluding object. Conversely, if the wavelength is hardly similar in size, then there is almost no wave diffraction perceivable at all. This relationship between the strength of wave diffraction and wavelength-obstacle-dimensions is conceptually illustrated in figure 2.11 when a wave is transmitted through a surface. A reflective example is provided in figure 2.9.

⁹Image source:http://cronodon.com/images/Single_slit_diffraction_2b.jpg

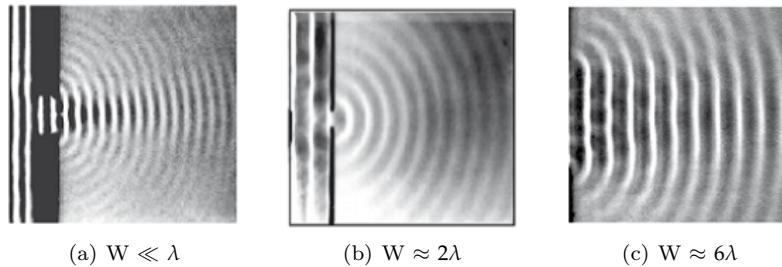


Abbildung 2.11: Illustration¹⁰ of how the effect of diffraction changes when a wave with wavelength λ propagates through a slit of width W .

In everyday's life, we can see the direct outcome of the effect of wave diffraction in form of structural colors. There are examples from nature such as the iridescent colors on various snake skins as well as artificial examples such as the colorful patterns notable when having a close look at an illuminated compact disc. All in common having having a surface made of very regular nanostructure which is diffracting an incident light. Such a nanostructure which exhibits a certain degree of regularity is also denoted as diffraction grating. More about this in section ??.

2.3 Stam's BRDF formulation

The theoretical foundation of this thesis is based on the pioneering work of J.Stam who derived in his paper about Diffraction Shader[Sta99] a BRDF which is modeling the effect of far field diffraction for various analytical anisotropic reflexion models, relying on the so called scalar wave theory of diffraction for which a wave is assumed to be a complex valued scalar.

It's noteworthy, that Stam's BRDF formulation does not take into account the polarization of the light. Fortunately, light sources like sunlight and light bulbs are unpolarized. The principal idea behind J. Stam's approach is illustrated in figure 2.12.

¹⁰Image taken from:<http://neutrino.ethz.ch/Vorlesung/FS2013/index.php/vorlesungsskript>, chapter 9, figure 9.14

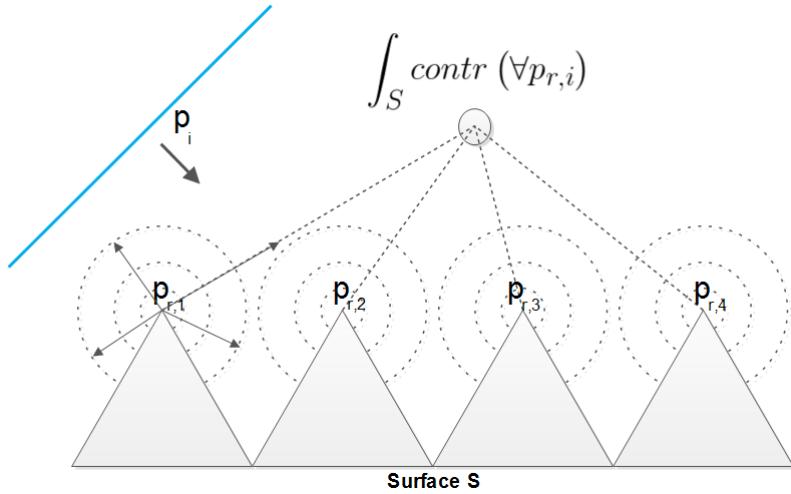


Abbildung 2.12: Illustration of idea behind Stam's approach: Integration over all secondary sources according to Huygen's principle resulting from an incident wave will give us an identity for the total contribution at a certain point in space.

An incident wave p_i from a light source encounters a surface, representing a diffraction grating. According to Huygen's Principle, at any point i on the grating, at which the incident wave meets the grating, a secondary, spherical wavelet $p_{r,i}$ will be emitted. A viewer, in the figure indicated by a gray circle, will perceive the superimposed contribution of all wavelets along the surface S (in the figure indicated by an integration symbol), which will directly follow the laws of wave interference. Therefore the resulting color which an observer sees is the final radiance at that point which itself is affected by stationary interference of all emitting secondary sources due to Huygen's principle.

A further assumption in Stam's Paper is, that the emanated waves from the source are stationary, which implies the wave is a superposition of independent monochromatic waves. This further implies that each wave is associated to a definite wavelength λ . However, directional light sources, such as sunlight fulfills this fact and since we are using these kinds of light sources for our simulations, Stam's model can be used for our modeling purposes.

The main idea of his model is to formulate a BRDF as the Fourier Transform applied on the given height field, representing a surface like shown in figure *fig : geometricsetup*. The classes of surfaces his model is able to support either exhibit a very regular structure or may be considered as a superposition of bumps forming a periodic like structure. Therefore, the surfaces he is dealing with can either be modeled by probabilistic distributions or have a direct analytical representation. Both cases allow him to derive an analytical solution for his BRDF model.

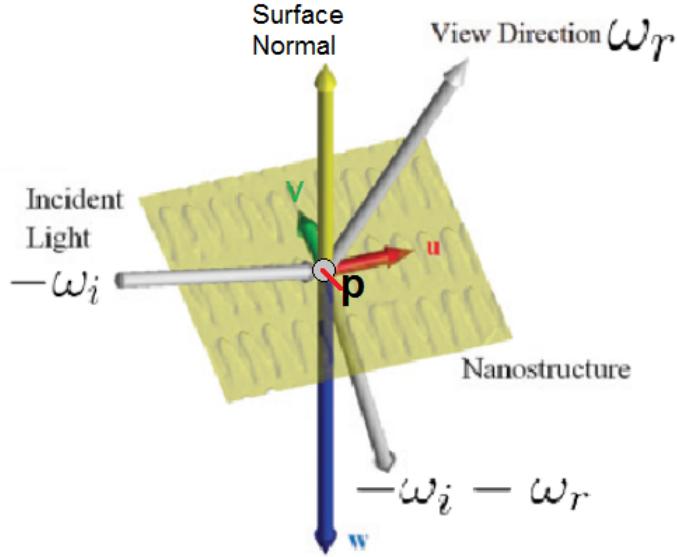


Abbildung 2.13: Illustration¹¹ of geometrical setup of Stam's approach where ω_i is a direction, pointing towards the light source, ω_r points towards the camera, n is the surface normal, (u, v, w) are the components of the vector $-\omega_i - \omega_r$.

The direction vector of the secondary wavelet can be computed by taking the difference between the incident and viewing direction like shown in equation 2.5:

$$(u, v, w) = -\omega_i - \omega_r \quad (2.5)$$

These coordinates will later be used in order to compute the total contribution of all secondary sources used in Stam's BRDF in equation 2.8. For simplification, let us introduce an auxiliary function Φ defined in equation 2.6, which models the phase of a wave from the provided height field.

$$\Phi(x, y) = \frac{2\pi}{\lambda} wh(x, y) \quad (2.6)$$

Then, any secondary wavelet p which is emitted off from the given surface will be equal:

$$p(x, y) = e^{i\Phi(x, y)} \quad (2.7)$$

using the idea presented for figure 2.12 and performing all mathematical steps shown in the appendix B, will lead us to the final BRDF representation, modeling the total contribution of all secondary sources reflected off the the provided surface h :

$$BRDF_\lambda(\omega_i, \omega_r) = \frac{k^2 F^2 G}{4\pi^2 A w^2} \langle |P(ku, kv)|^2 \rangle \quad (2.8)$$

where F denotes the Fresnel coefficient and G is the so called geometry term which is equal:

$$G = \frac{(1 + \omega_i \cdot \omega_r)^2}{\cos(\theta_i) \cos(\theta_r)} \quad (2.9)$$

¹¹Modified image which originally has been taken from D.H. Dhillon's poster[DSD14].

One last word about the term Fourier Transform Stam uses in his derivation: Conventionally, following the definitions in Mathematics of the Fourier Transformation, we are dealing with the inverse Fourier Transformation. However, especially in Electrical Engineering, it is quite common to define this inverse Fourier Transformation by the Fourier Transformation. The reason behind this lies in the fact that we simply could substitute the minus sign like the following equation 2.10:

$$\begin{aligned}\mathcal{F}_{FT}\{f\}(w) &= \int_{\mathbb{R}^n} f(x)e^{-iwt} dt \\ &= \int_{\mathbb{R}^n} f(x)e^{i\hat{w}t} dt \\ &= \mathcal{F}_{FT}^{-1}\{f\}(w)\end{aligned}\quad (2.10)$$

where \hat{w} is equal $-w$.

The height fields we are dealing with in this work are, however, natural gratings, containing a complex shaped nano-structure and hence far apart from being very regularly aligned. The reason why Stam's approach in its current form is not suitable for our purpose is twofold: First his approach does not capture the complexity of natural gratings accurately well enough when relying on his statistical approaches and secondly it is way too slow in order to be usable for interactive rendering since his BRDF needs an evaluation of a Fourier Transform for every directional changement.

In the following a brief comparison between Stam's¹² and our final approach using two different kinds of gratings, a synthetic, regularly aligned grating and a natural, complex structured grating. These gratings are shown in figure *fig : stamegratings*:

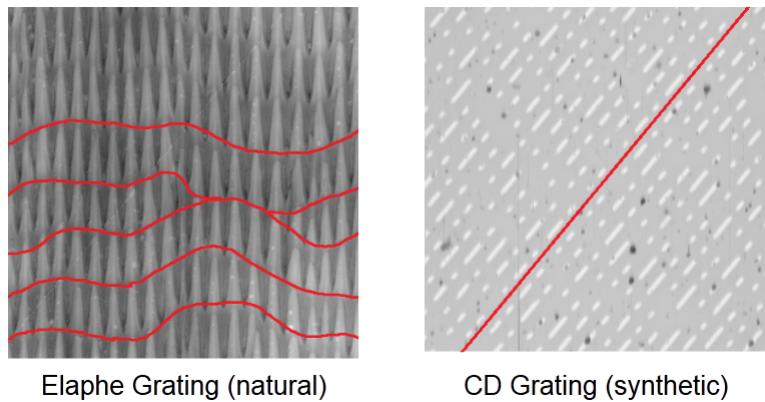


Abbildung 2.14: Alignment of nanostructures in diffraction gratings. On the left a complex, natural grating of the Elaphe snake species and on the right a synthetic, very regularly aligned grating of a CD.

¹²A reference implementation of Stam's Diffraction Shader[Sta99] is provided by Nvidia's GPU Gems at http://http.developer.nvidia.com/GPUGems/gpugems_08.html

Figure 2.15 shows an example of a case where Stam's approach performs well. Considering the red-line in the figure we notice that the nano-scaled structures of a compact disc are very regularly aligned along the surface. Tracks of a CD are uniformly spaced and bumps along a track are distributed according to the poisson distribution¹³. All angles of the diffraction pattern look the same as in the images produced by our approach.

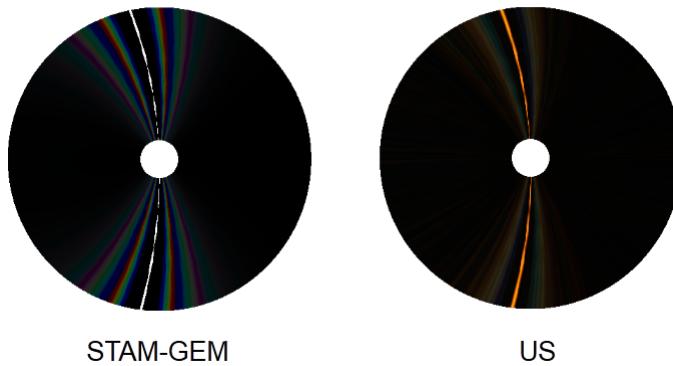


Abbildung 2.15: Comparission of our approach against a reference implementation of Stam's provided by Nvidia Gem. For synthetic diffraction gratings, which have a very regular structure, Stam's approach is doing well. All angles of the diffraction pattern look the same as in the images produced by our approach.

Figure 2.16 shows an attempt to reproduce real structural colors on the skin of the Xenopeltis snake using our method and comparing it to Stam's approach. Even the results of Stam might look appropriate, there are some differences notable such missing colors close to specular regions, or such as the color distribution which is rather discrete in Stam's approach. Furthermore, Stam's approach has at some places color contribution, where it should not have.

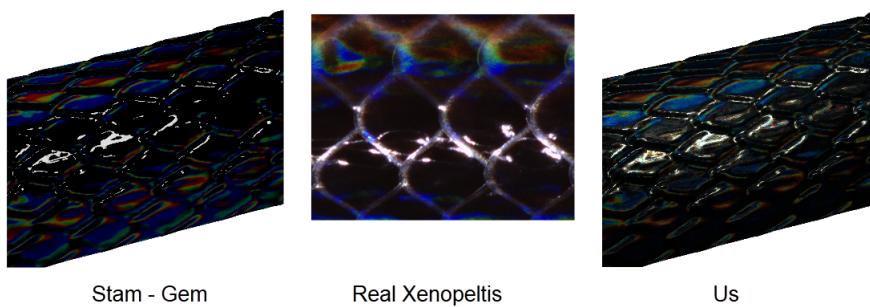


Abbildung 2.16: Comparission of our approach(on the right) against a reference implementation of Stam's(on the left) provided by Nvidia Gem by trying to render a reproduction of a real Elaphe skin (center) under similar lightning and viewing conditions. For natural diffraction gratings, which have a rather complex structure, Stam's approach is doing rather bad.

¹³See http://en.wikipedia.org/wiki/Poisson_distribution

In the next chapter we are going to adapt Stam's BRDF model such that it will be able to handle the kind of surfaces we are dealing with and even will have a runtime complexity which allows to perform interactive rendering.

Kapitel 3

Derivations

3.1 Problem Statement and Challenges

The goal of this thesis is to perform a physically accurate and interactive simulation of structural colors production like shown in figure 3.2, which we can see whenever a light source is diffracted on a natural grating. For this purpose we need the to be provided by the following input data as shown in figure 3.1:

- A mesh representing a snake surface¹ with associated texture coordinates as shown in figure 3.1(a).
- A natural diffraction grating represented as a height field, its maximum height and its pixel-width-correspondence².
- A vectorfield which describes how fingers on a provided surface of the nanostructure are aligned as shown in figure 3.1(c).

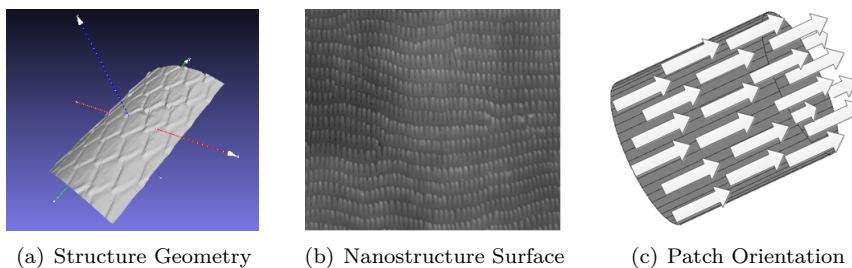


Abbildung 3.1: Input for our simulation

We want to rely on the integral equation 2.8 derived by J. Stam in his paper [Sta99] about diffraction shaders. This equation formualtes a BRDF modeling

¹Which is in our simulation an actual reconstruction of a real snake skin. These measurements are provided by the Laboratory of Artificial and Natural Evolution at Geneva. See their website: www.lanevol.org.

²Since the nanostructure is stored as a grayscale image, we need a scale telling us what length and height one pixel corresponds to in this provided image.

the effect of diffraction under the assumption that a given grating can either be formulated as an analytical function or its structure is simple enough being modeled relying on statistical methods. These assumptions guarantee that 2.8 has an explicit solution. However, the complexity of a biological nanostructures cannot sufficiently and accurately modeled simply using statistical methods. This is why interactive computation at high resolution becomes a hard task, since we cannot evaluate the given integral equation on the fly. Therefore, we have to adapt Stam's equation such that we are able to perform interactive rendering using explicitly provided height fields.

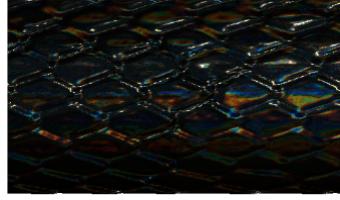


Abbildung 3.2: Output: Rendered Structural Colors

3.2 Rendering Equation

As already discussed in the theoretical background chapter, colors are associated to radiance. Since we are starting with Stam's BRDF³ formulation but want to perform a simulation rendering structural colors, we have to reformulate this BRDF equation such that we will end up with an identity of the reflected spectral radiance. This is where the rendering equation comes into play. Lets assume we have given an incoming light source with solid angle ω_i and θ_i is its angle of incidence, ω_r is the solid angle for the reflected light. Further let λ denote the wavelength and Ω is the hemisphere we of integration for the incoming light. Then, we are able to formulate a BRDF by using its definition 2.3:

$$\begin{aligned}
 f_r(\omega_i, \omega_r) &= \frac{dL_r(\omega_r)}{L_i(\omega_i)\cos(\theta_i)d\omega_i} \\
 \Rightarrow f_r(\omega_i, \omega_r)L_i(\omega_i)\cos(\theta_i)d\omega_i &= dL_r(\omega_r) \\
 \Rightarrow \int_{\Omega} f_r(\omega_i, \omega_r)L_i(\omega_i)\cos(\theta_i)d\omega_i &= \int_{\Omega} dL_r(\omega_r) \\
 \Rightarrow L_r(\omega_r) &= \int_{\Omega} f_r(\omega_i, \omega_r)L_i(\omega_i)\cos(\theta_i)d\omega_i \quad (3.1)
 \end{aligned}$$

The last equation is the so called rendering equation . We assume that our incident light is a directional, unpolarized light source ?? like sunlight and therefore its radiance is given as

$$L_{\lambda}(\omega) = I(\lambda)\delta(\omega - \omega_i) \quad (3.2)$$

where $I(\lambda)$ is the intensity of the relative spectral power for the wavelength λ . Since all light rays are parallel, whenever we are provided by a directional

³Remember that a BRDF is the portion of a incident light source reflected off a given surface towards a specified viewing direction

light source and we can think of radiance as a measure of the light emitted from a particular surface location into a particular direction, above's radiance identity will follow immediately. By plugging the identity 3.2 into our current rendering equation 3.1, we will get:

$$\begin{aligned} L_\lambda(\omega_r) &= \int_{\Omega} BRDF_\lambda(\omega_i, \omega_r) L_\lambda(\omega_i) \cos(\theta_i) d\omega_i \\ &= BRDF_\lambda(\omega_i, \omega_r) I(\lambda) \cos(\theta_i) \end{aligned} \quad (3.3)$$

where $L_\lambda(\omega_i)$ is the incoming radiance and $L_\lambda(\omega_r)$ is the radiance reflected by the given surface. Note that the integral in equation 3.3 vanishes since $\delta(\omega - \omega_i)$ is only equal one if and only if $\omega = \omega_i$.

3.3 Adaption of Stam's BRDF discrete height fields

We are going to use Stam's main derivation (2.8) for the $BRDF(\omega_i, \omega_r)$ in 3.3 by applying the fact that the wavenumber is equal $k = \frac{2\pi}{\lambda}$:

$$\begin{aligned} BRDF(\omega_i, \omega_r) &= \frac{k^2 F^2 G}{4\pi^2 A w^2} \langle |P(ku, kv)|^2 \rangle \\ &= \frac{4\pi^2 F^2 G}{4\pi^2 A \lambda^2 w^2} \langle |P(ku, kv)|^2 \rangle \\ &= \frac{F^2 G}{A \lambda^2 w^2} \left\langle \left| P\left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda}\right) \right|^2 \right\rangle \end{aligned} \quad (3.4)$$

Going back to equation 3.3 and plugging equation 3.4 into it, using the definition of equation 2.9 and the equation 3.12 for ω we will get the following:

$$\begin{aligned} L_\lambda(\omega_r) &= \frac{F^2 (1 + \omega_i \cdot \omega_r)^2}{A \lambda^2 \cos(\theta_i) \cos(\theta_r) \omega^2} \left\langle \left| P\left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda}\right) \right|^2 \right\rangle \cos(\theta_i) I(\lambda) \\ &= I(\lambda) \frac{F^2 (1 + \omega_i \cdot \omega_r)^2}{\lambda^2 A \omega^2 \cos(\theta_r)} \left\langle \left| P\left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda}\right) \right|^2 \right\rangle \end{aligned} \quad (3.5)$$

Note that the Fresnel term F is actually a function of (w_i, w_r) , but in order to keep the equations simple, we omitted its arguments. So far we just plugged Stam's BRDF identity into the rendering equation and hence have not significantly deviated from his formulation. Keep in mind that P denotes the Fourier transform of the provided height field which depends on the viewing and incidence light direction. Thus this Fourier Transform has to be recomputed for every direction which will slow down the whole computation quite a lot⁴. One particular strategy to solve this issue is to approximate P by the Discrete Fourier Transform (DFT)⁵ and separate its computation such that terms for many

⁴Even a fast variant of computation the Fourier Transform has a runtime complexity of $O(N \log N)$ where N is the number of samples.

⁵See appendix A for further information about different kinds of Fourier transformations.

directions can be precomputed and then later retrieved by look ups. The approximation of P happens in two steps: First we approximate the Fourier Transform by the Discrete Time Fourier Transform (DTFT) and then, afterwards, we approximate the DTFT by the DFT. For further about basics of signal processing and Fourier Transformations please consult the appendix A.

3.4 Reproduce FT by DTFT

In the previous section, we have found an identity for the reflected spectral radiance $L_\lambda(\omega_r)$ when using Stam's BRDF for a given input height field. However, the derived expression in equation 3.5 requires to evaluate the Fourier Transform of our height field⁶ for every direction. In this section we explain how to approximate the FT by the DTFT and apply it to our previous derivations. Figure 3.3 graphically shows how to obtain the DTFT from the FT for a one dimensional signal⁷

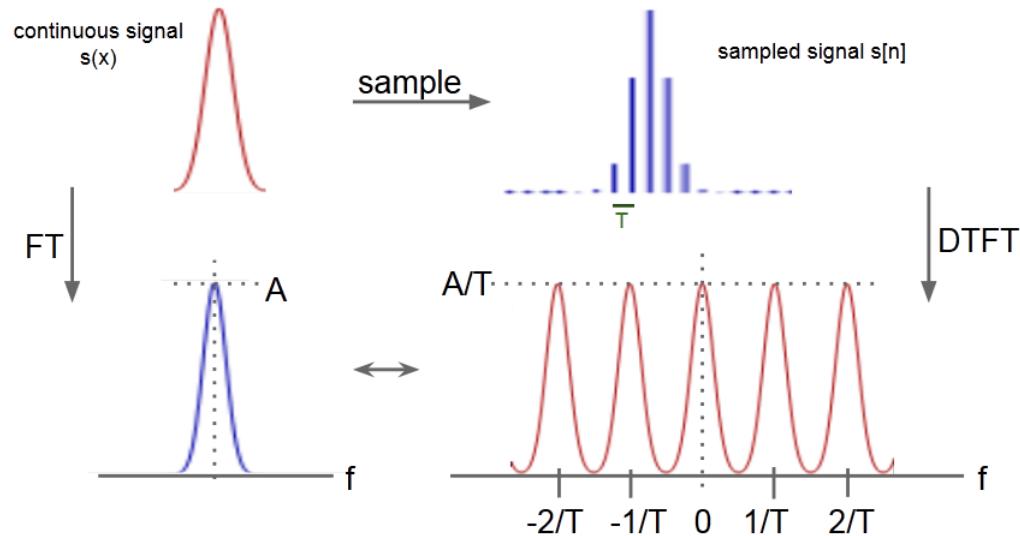


Abbildung 3.3: Illustration of how to approximate the analytical Fourier Transform (FT)⁸ of a given continuous signal by a Discrete Time Fourier Transform (DTFT).

The first step is to uniformly discretize the given signal since computers are working finite, discrete arithmetic. We rely on the Nyquist–Shannon sampling

⁶actually it requires the computation of the inverse Fourier Transform of a transformed version of the given heightfield, the function $p(x,y)$ defined in equation 2.7.

⁷For our case we are dealing with a two dimensional, spatial signal, the given height field. Nevertheless, without any constraints of generality, the explained approach applies to multi dimensional problems.

⁸Images of function plots taken from http://en.wikipedia.org/wiki/Discrete_Fourier_transform and are modified.

theorem tells us how dense we have to sample a given signal $s(x)$ such that can be reconstructed its sampled version $\hat{s}[n]$ ⁹. In particular, a sampled version according to the Nyquist–Shannon sampling theorem will have the same Fourier Transform as its original singal. The sampling theorem states that if f_{max} denotes the highest frequency of $s(x)$, then, it has to be sampled by a rate of f_s with $2f_{max} \leq f_s$ in order to be reconstructable. By convention $T = \frac{1}{f_s}$ represent the interval length between two samples.

Next, we apply the Fourier Tranformation operator on the discretized singal \hat{s} which gives us the following expression:

$$\begin{aligned}\mathcal{F}_{FT}\{\hat{s}\}(w) &= \int_{\mathbb{R}} \hat{s}[n] e^{-iwx} dx \\ &= \int_{\mathbb{R}} mask(x) s(x) e^{-iwx} dx \\ &= T \sum_{x=-\infty}^{\infty} \hat{s}[x] e^{-iwx} \\ &= T \mathcal{F}_{DTFT}\{s\}(w)\end{aligned}\tag{3.6}$$

Equation 3.6 tells us that if \hat{s} is sufficiently sampled, then its DTFT corresponds to the FT of $s(x)$. Notice that the resulting DTFT from the sampled signal has a height of $\frac{A}{T}$ where A is the height of the FT of s and thus is a scaled version of the FT.

Using the insight gained by equation 3.6 allows us to further simplify equation 3.5:

$$\begin{aligned}L_{\lambda}(\omega_r) &= I(\lambda) \frac{F^2(1 + \omega_i \cdot \omega_r)^2}{\lambda^2 A w^2 \cos(\theta_r)} \left\langle \left| P\left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda}\right) \right|^2 \right\rangle \\ &= I(\lambda) \frac{F^2(1 + \omega_i \cdot \omega_r)^2}{\lambda^2 A w^2 \cos(\theta_r)} \left\langle \left| T^2 P_{dtft}\left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda}\right) \right|^2 \right\rangle\end{aligned}\tag{3.7}$$

Where P_{dtft} is a substitute for $\mathcal{F}_{DTFT}\{s\}(w)$. Furthermore T the sampling distance for the discretization of $p(x, y)$ assuming equal and uniform sampling in both dimensions x and y .

3.5 Special Coherence and Windowing

3.6 Reproduce DTFT by DFT

Practically, we cannot compute the DTFT A.3 numerically due to finite computer arithmetic, since w is a continuous function for the DTFT. The DFT A.4 of

⁹n denotes the number of samples.

a discrete height field patch is equivalent to the DTFT of an infinitely periodic function consisting of replicas of the same discrete patch. By windowing with a window function that is zero outside the central replica, the convolution of either the DFT or the DTFT of height field with the Fourier Transform of the window becomes equivalent.

Let $window_g$ denote the gaussian window with $4\sigma_s \mu m$ where $\sigma_f = \frac{1}{2\pi\sigma_s}$ let us further substitute $\mathbf{t}(\mathbf{x}, \mathbf{y}) = i^n h(x, y)^n$

$$\mathcal{F}_{dtft}^{-1}\{\mathbf{t}\}(u, v) = \mathcal{F}_{fft}^{-1}\{\mathbf{t}\}(u, v)window_g(\sigma_f) \quad (3.8)$$

Therefore we can deduce the following expression from this:

$$\begin{aligned} \mathcal{F}_{dtft}^{-1}\{\mathbf{t}\}(u, v) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_{fft}^{-1}\{\mathbf{t}\}(w_u, w_v) \phi(u - w_u, v - w_v) dw_u dw_v \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_i \sum_j F_{fft}^{-1}\{\mathbf{t}\}(w_u, w_v) \\ &\quad \delta(w_u - w_i, w_v - w_j) \phi(u - w_u, v - w_v) dw_u dw_v \\ &= \sum_i \sum_j \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_{fft}^{-1}\{\mathbf{t}\}(w_u, w_v) \\ &\quad \delta(w_u - w_i, w_v - w_j) \phi(u - w_u, v - w_v) dw_u dw_v \\ &= \sum_i \sum_j F_{fft}^{-1}\{\mathbf{t}\}(w_u, w_v) \phi(u - w_u, v - w_v) \end{aligned} \quad (3.9)$$

where

$$\phi(x, y) = \pi e^{-\frac{x^2+y^2}{2\sigma_f^2}} \quad (3.10)$$

3.7 Relative Reflectance

In this section we are going to explain how to scale our BRDF formulation such that all of its possible output values are mapped into the range $[0, 1]$. Such a relative reflectance formulation will ease our life for later rendering purposes since usually color values are within the range $[0, 1]$, too. Furthermore, this will allow us to properly blend the resulting illumination caused by diffraction with a texture map.

Let us examine what $L_{\lambda}(\omega_r)$ will be for $\omega_r = \omega_0 := (0, 0, *)$ i.e. specular reflection case, denoted as $L_{\lambda}^{spec}(\omega_0)$. When we know the expression for $L_{\lambda}^{spec}(\omega_0)$ we would be able to compute the relative reflected radiance for our problem 3.5 by simply taking the fraction between $L_{\lambda}(\omega_r)$ and $L_{\lambda}^{spec}(\omega_0)$ which is denoted by:

$$\rho_{\lambda}(\omega_i, \omega_r) = \frac{L_{\lambda}(\omega_r)}{L_{\lambda}^{spec}(\omega_0)} \quad (3.11)$$

Going back to the definition 2.5 of $(u, v, w) = -\omega_i - \omega_r$ and using spherical coordinates C.2, we get for w the following identity

$$\begin{aligned}
w &= -\omega_i - \omega_r \\
&= -(\omega_i + \omega_r) \\
&= -(\cos(\theta_i) + \cos(\theta_r))
\end{aligned} \tag{3.12}$$

and therefore w^2 is equal $(\cos(\theta_i) + \cos(\theta_r))^2$.

But first, let us derive the following expression:

$$\begin{aligned}
L_{\lambda}^{spec}(\omega_0) &= I(\lambda) \frac{F(\omega_0, \omega_0)^2 (1 + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix})^2}{\lambda^2 A (\cos(0) + \cos(0))^2 \cos(0)} \langle \left| T_0^2 P_{dtft}(0, 0) \right|^2 \rangle \\
&= I(\lambda) \frac{F(\omega_0, \omega_0)^2 (1+1)^2}{\lambda^2 A (1+1)^2 1} \left| T_0^2 N_{sample} \right|^2 \\
&= I(\lambda) \frac{F(\omega_0, \omega_0)^2}{\lambda^2 A} \left| T_0^2 N_{sample} \right|^2
\end{aligned} \tag{3.13}$$

Where $N_{samples}$ is the number of samples of the DTFT A.3. Thus, we can plug our last derived expression 3.13 into the definition for the relative reflectance radiance 3.11 in the direction w_r and will get:

$$\begin{aligned}
\rho_{\lambda}(\omega_i, \omega_r) &= \frac{L_{\lambda}(\omega_r)}{L_{\lambda}^{spec}(\omega_0)} \\
&= \frac{I(\lambda) \frac{F(\omega_i, \omega_r)^2 (1+\omega_i \cdot \omega_r)^2}{\lambda^2 A (\cos(\theta_i) + \cos(\theta_r))^2 \cos(\theta_r)} \langle \left| T_0^2 P_{dtft}(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda}) \right|^2 \rangle}{I(\lambda) \frac{F(\omega_0, \omega_0)^2}{\lambda^2 A} \left| T_0^2 N_{sample} \right|^2} \\
&= \frac{F^2(\omega_i, \omega_r) (1 + \omega_i \cdot \omega_r)^2}{F^2(\omega_0, \omega_0) (\cos(\theta_i) + \cos(\theta_r))^2 \cos(\theta_r)} \langle \left| \frac{P_{dtft}(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda})}{N_{samples}} \right|^2 \rangle
\end{aligned} \tag{3.14}$$

For simplification and a better overview, let us introduce the following expression, the so called gain-factor:

$$C(\omega_i, \omega_r) = \frac{F^2(\omega_i, \omega_r) (1 + \omega_i \cdot \omega_r)^2}{F^2(\omega_0, \omega_0) (\cos(\theta_i) + \cos(\theta_r))^2 \cos(\theta_r) N_{samples}^2} \tag{3.15}$$

Using this substitute, we will end up with the following expression for the relative reflectance radiance from equation 3.14:

$$\rho_{\lambda}(\omega_i, \omega_r) = C(\omega_i, \omega_r) \langle \left| P_{dtft}(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda}) \right|^2 \rangle \tag{3.16}$$

Using the previous definition for the relative reflectance radiance 3.11:

$$\rho_{\lambda}(\omega_i, \omega_r) = \frac{L_{\lambda}(\omega_r)}{L_{\lambda}^{spec}(\omega_0)} \tag{3.17}$$

Which we can rearrange to the expression:

$$L_\lambda(\omega_r) = \rho_\lambda(\omega_i, \omega_r) L_\lambda^{spec}(\omega_0) \quad (3.18)$$

Let us choose $L_\lambda^{spec}(\omega_0) = S(\lambda)$ such that it has the same profile as the relative spectral power distribution of CIE Standard Illuminant $D65$ discussed in ???. Furthermore, when integrating over λ for a specular surface, we should get CIE_{XYZ} values corresponding to the white point for $D65$. The corresponding tristimulus values using CIE colormatching functions 2.4 for the CIE_{XYZ} values look like:

$$\begin{aligned} X &= \int_{\lambda} L_\lambda(\omega_r) \bar{x}(\lambda) d\lambda \\ Y &= \int_{\lambda} L_\lambda(\omega_r) \bar{y}(\lambda) d\lambda \\ Z &= \int_{\lambda} L_\lambda(\omega_r) \bar{z}(\lambda) d\lambda \end{aligned} \quad (3.19)$$

where \bar{x} , \bar{y} , \bar{z} are the color matching functions. Using our last finding 3.18 for $L_\lambda(\omega_r)$ with the definition for the tristimulus values 3.19, we can actually derive an expression for computing the colors for our initial BRDF formula 3.1. Without any loss of generality it satisfies to derive an explicit expression for just one tristimulus term, for example X . Since The other have a similar formulation, except the we have to replace all X with Y or Z respectively. Therefore, we get:

$$\begin{aligned} X &= \int_{\lambda} L_\lambda(\omega_r) \bar{x}(\lambda) d\lambda \\ &= \int_{\lambda} \rho_\lambda(\omega_i, \omega_r) L_\lambda^{spec}(\omega_0) \bar{x}(\lambda) d\lambda \\ &= \int_{\lambda} \rho_\lambda(\omega_i, \omega_r) S(\lambda) \bar{x}(\lambda) d\lambda \\ &= \int_{\lambda} C(\omega_i, \omega_r) \left\langle \left| P_{dtft} \left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda} \right) \right|^2 \right\rangle S(\lambda) \bar{x}(\lambda) d\lambda \\ &= C(\omega_i, \omega_r) \int_{\lambda} \left\langle \left| P_{dtft} \left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda} \right) \right|^2 \right\rangle S(\lambda) \bar{x}(\lambda) d\lambda \\ &= C(\omega_i, \omega_r) \int_{\lambda} \left\langle \left| P_{dtft} \left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda} \right) \right|^2 \right\rangle S_x(\lambda) d\lambda \end{aligned} \quad (3.20)$$

Where we used the definition $S_x(\lambda) \bar{x}(\lambda)$ in the last step.

3.8 Optimization using Taylor Series

In this section, we will deliver an approximation for the inverse Fourier Transformation of Stam's auxiliary function $p(x, y)$. This derivation will rely on the definition of Taylor Series expansion A.8. Further, we will provide an error bound for our approximation approach for a given number of iterations. Last, we will extend our current BRDF formula by the findings derived within this section.

Given $p(x, y) = e^{ikwh(x,y)}$ form Stam's Paper 2.3 where $h(x, y)$ is a given height field. Let be y real or even complex value, and lets consider the power series for the the exponential function

$$e^t = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \dots = \sum_{n=0}^{\infty} \frac{t^n}{n!} \quad (3.21)$$

Let us define

$$t = t(x, y) = ikwh(x, y) \quad (3.22)$$

where i is the imaginary number. For simplification, let us denote $h(x, y)$ as h . Then it follows by our previous stated identities:

$$\begin{aligned} e^t &= 1 + (ikwh) + \frac{1}{2!}(ikwh)^2 + \frac{1}{3!}(ikwh)^3 + \dots \\ &= \sum_{n=0}^{\infty} \frac{(ikwh)^n}{n!}. \end{aligned} \quad (3.23)$$

Hence it holds $p(x, y) = \sum_{n=0}^{\infty} \frac{(ikwh(x,y))^n}{n!}$. Let us now compute the Fourier Transformation of $p(x,y)$ form above:

$$\begin{aligned} \mathcal{F}\{p\}(u, v) &= \mathcal{F}\left\{\sum_{n=0}^{\infty} \frac{(ikwh)^n}{n!}\right\}(u, v) \\ &\stackrel{\text{= }\mathcal{F}\text{ lin Operator}}{=} \sum_{n=0}^{\infty} \mathcal{F}\left\{\frac{(ikwh)^n}{n!}\right\}(u, v) \\ &= \sum_{n=0}^{\infty} \frac{(ikw)^n}{n!} \mathcal{F}\{h^n\}(u, v) \end{aligned} \quad (3.24)$$

Therefore it follows: $P(\alpha, \beta) = \sum_{n=0}^{\infty} \frac{(ikw)^n}{n!} \mathcal{F}\{h^n\}(\alpha, \beta)$ for which $\mathcal{F}_{FT}\{h^n\}(u, v)$. Next we are going to look for an $N \in \mathbb{N}$ such that

$$\sum_{n=0}^N \frac{(ikwh)^n}{n!} \mathcal{F}\{h^n\}(\alpha, \beta) \approx P(\alpha, \beta) \quad (3.25)$$

is a good approximation. But first the following two facts have to be proven:

1. Show that there exist such an $N \in \mathbb{N}$ s.t the approximation holds true.
2. Find a value for B s.t. this approximation is below a certain error bound, for example machine precision ϵ .

Proof Sketch of 1.

By the **ratio test** (see [1]) It is possible to show that the series $\sum_{n=0}^{\infty} \frac{(ikwh)^n}{n!} \mathcal{F}\{h^n\}(\alpha, \beta)$ converges absolutely:

Proof: Consider $\sum_{k=0}^{\infty} \frac{y^n}{n!}$ where $a_k = \frac{y^k}{k!}$. By applying the definition of the ratio test for this series it follows:

$$\forall y : \limsup_{k \rightarrow \infty} \left| \frac{a_{k+1}}{a_k} \right| = \limsup_{k \rightarrow \infty} \frac{y}{k+1} = 0 \quad (3.26)$$

Thus this series converges absolutely, no matter what value we will pick for y .

Part 2: Find such an N

Let $f(x) = e^x$. We can formulate its Taylor-Series, stated above. Let $P_n(x)$ denote the n-th Taylor polynom,

$$P_n(x) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k \quad (3.27)$$

where a is our developing point (here a is equal zero).

We can define the error of the n-th Taylor polynom to be $E_n(x) = f(x) - P_n(x)$. the error of the n-th Taylor polynom is difference between the value of the function and the Taylor polynomial This directly implies $|E_n(x)| = |f(x) - P_n(x)|$. By using the Lagrangian Error Bound it follows:

$$|E_n(x)| \leq \frac{M}{(n+1)!} |x-a|^{n+1} \quad (3.28)$$

with $a = 0$, where M is some value satisfying $|f^{(n+1)}(x)| \leq M$ on the interval $I = [a, x]$. Since we are interested in an upper bound of the error and since a is known, we can reformulate the interval as $I = [0, x_{max}]$, where

$$x_{max} = \|i\| k_{max} w_{max} h_{max} \quad (3.29)$$

We are interested in computing an error bound for $e^{ikwh(x,y)}$. Assuming the following parameters and facts used within Stam's Paper:

- Height of bump: 0.15micro meters
- Width of a bump: 0.5micro meters
- Length of a bump: 1micro meters
- $k = \frac{2\pi}{\lambda}$ is the wavenumber, $\lambda \in [\lambda_{min}, \lambda_{max}]$ and thus $k_{max} = \frac{2\pi}{\lambda_{min}}$. Since $(u, v, w) = -\omega_i - \omega_r$ and both are unit direction vectors, each component can have a value in range [-2, 2].
- for simplification, assume $[\lambda_{min}, \lambda_{max}] = [400nm, 700nm]$.

We get:

$$\begin{aligned} x_{max} &= \|i\| * k_{max} * w_{max} * h_{max} \\ &= k_{max} * w_{max} * h_{max} \\ &= 2 * \left(\frac{2\pi}{4 * 10^{-7} m} \right) * 1.5 * 10^{-7} \\ &= 1.5\pi \end{aligned} \quad (3.30)$$

and it follows for our interval $I = [0, 1.5\pi]$.

Next we are going to find the value for M . Since the exponential function is monotonically growing (on the interval I) and the derivative of the **exp** function is the exponential function itself, we can find such an M :

$$\begin{aligned} M &= e^{x_{max}} \\ &= \exp(1.5\pi) \end{aligned}$$

and $|f^{(n+1)}(x)| \leq M$ holds. With

$$\begin{aligned} |E_n(x_{max})| &\leq \frac{M}{(n+1)!} |x_{max} - a|^{n+1} \\ &= \frac{\exp(1.5\pi) * (1.5\pi)^{n+1}}{(n+1)!} \end{aligned} \quad (3.31)$$

we now can find a value of n for a given bound, i.e. we can find an value of $N \in \mathbb{N}$ s.t. $\frac{\exp(1.5\pi) * (1.5\pi)^{N+1}}{(N+1)!} \leq \epsilon$. With Octave/Matlab we can see:

- if $N=20$ then $\epsilon \approx 2.9950 * 10^{-4}$
- if $N=25$ then $\epsilon \approx 8.8150 * 10^{-8}$
- if $N=30$ then $\epsilon \approx 1.0050 * 10^{-11}$

With this approach we have that $\sum_{n=0}^{25} \frac{(ikwh)^n}{n!} \mathcal{F}\{h^n\}(\alpha, \beta)$ is an approximation of $P(u, v)$ with error $\epsilon \approx 8.8150 * 10^{-8}$. This means we can precompute 25 Fourier Transformations in order to approximate $P(u, v)$ having an error $\epsilon \approx 8.8150 * 10^{-8}$.

Using now our approximation for $P_{dtft} = \mathcal{F}^{-1}\{p\}(u, v)$ for the tristimulus value X , we will get:

$$\begin{aligned} X &= C(w_i, w_r) \int_{\lambda} \left\langle \left| P_{dtft} \left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda} \right) \right|^2 \right\rangle S_x(\lambda) d\lambda \\ &= C(w_i, w_r) \int_{\lambda} \left\langle \sum_{n=0}^N \frac{(wk)^n}{n!} \mathcal{F}^{-1}\{i^n h^n\} \left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda} \right) \right\rangle^2 S_x(\lambda) d\lambda \end{aligned} \quad (3.32)$$

3.9 Spectral Rendering

As the last step of our series of derivations, we plug all our findings together to one big equation in order to compute the color for each pixel on our mesh in the CIE_{XYZ} colorspace. For any given height-field $h(x, y)$ representing a small patch of a nano structure of a surface and the direction vectors w_s and w_r from figure 2.13 the resulting color caused by the effect of diffraction can be computed like: Let

$$P_{\lambda}(u, v) = F_{fft}^{-1}\{i^n h^n\} \left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda} \right) \quad (3.33)$$

Then our final expression using our previous derivations will look like:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = C(\omega_i, \omega_r) \int_{\lambda} \sum_{n=0}^N \frac{(wk)^n}{n!} \sum_{(r,s) \in \mathcal{N}_1(u,v)} |P_{\lambda}(u - w_r, v - w_s)|^2 \phi(u - w_r, v - w_s) \begin{pmatrix} S_x(\lambda) \\ S_y(\lambda) \\ S_z(\lambda) \end{pmatrix} d\lambda \quad (3.34)$$

where $\phi(x, y) = \pi e^{-\frac{x^2+y^2}{2\sigma_f^2}}$ is the Gaussian window 3.6.

3.10 Alternative Approach

3.10.1 PQ factors

In this section we are presenting an alternative approach to the previous Gaussian window approach 3.6 in order to solve the issue working with *DTFT* instead the *DFT*. We assume, that a given surface S is covered by a number of replicas of a provided representative surface patch f . In a simplified, one dimensional scenario, mathematically speaking, f is assumed to be a periodic function, i.e. $\forall x \in \mathbb{R} : f(x) = f(x + nT)$, where T is its period and $n \in \mathbb{N}_0$. Thus, the surfaces can be written formally as:

$$S(x) = \sum_{n=0}^N f(x + nT) \quad (3.35)$$

What we are looking for is an identity for the inverse Fourier transform of our surface S , required in order to simplify the (X, Y, Z) colors from 3.32:

$$\begin{aligned} \mathcal{F}^{-1}\{S\}(w) &= \int f(x) e^{iwx} dx \\ &= \int_{-\infty}^{\infty} \sum_{n=0}^N f(x + nT) e^{iwx} dx \\ &= \sum_{n=0}^N \int_{-\infty}^{\infty} f(x + nT) e^{iwx} dx \end{aligned} \quad (3.36)$$

Next, apply the following substitution $x + nT = y$ which will lead us to:

$$\begin{aligned} x &= y - nT \\ dx &= dy \end{aligned} \quad (3.37)$$

Plugging this substitution back into equation 3.36 we will get:

$$\begin{aligned}
\mathcal{F}^{-1}\{S\}(w) &= \sum_{n=0}^N \int_{-\infty}^{\infty} f(x+nT) e^{iwx} dx \\
&= \sum_{n=0}^N \int_{-\infty}^{\infty} f(y) e^{iw(y-nT)} dy \\
&= \sum_{n=0}^N e^{-iwnT} \int_{-\infty}^{\infty} f(y) e^{iwy} dy \\
&= \sum_{n=0}^N e^{-iwnT} \mathcal{F}^{-1}\{f\}(w) \\
&= \mathcal{F}^{-1}\{f\}(w) \sum_{n=0}^N e^{-iwnT}
\end{aligned} \tag{3.38}$$

We used the fact that the exponential term e^{-iwnT} is a constant factor when integrating along dy and the identity for the inverse Fourier transform of the function f . Next, let us examine the series $\sum_{n=0}^N e^{-iwnT}$ closer:

$$\begin{aligned}
\sum_{n=0}^N e^{-uwnT} &= \sum_{n=0}^N (e^{-uwT})^n \\
&= \frac{1 - e^{iwT(N+1)}}{1 - e^{-iwT}}
\end{aligned} \tag{3.39}$$

We recognize the geometric series identity for the left-hand-side of equation 3.39. Since our series is bounded, we can simplify the right-hand-side of equation 3.39.

Note that e^{-ix} is a complex number. Every complex number can be written in its polar form, i.e.

$$e^{-ix} = \cos(x) + i\sin(x) \tag{3.40}$$

Using the following trigonometric identities

$$\begin{aligned}
\cos(-x) &= \cos(x) \\
\sin(-x) &= -\sin(x)
\end{aligned} \tag{3.41}$$

combined with 3.40 we can simplify the series 3.39 even further to:

$$\frac{1 - e^{iwT(N+1)}}{1 - e^{-iwT}} = \frac{1 - \cos(wT(N+1)) + i\sin(wT(N+1))}{1 - \cos(wT) + i\sin(wT)} \tag{3.42}$$

Equation 3.42 is still a complex number, denoted as $(p + iq)$. Generally, every complex number can be written as a fraction of two complex numbers. This implies that the complex number $(p + iq)$ can be written as $(p + iq) = \frac{(a+ib)}{(c+id)}$ for any $(a + ib), (c + id) \neq 0$. Let us use the following substitutions:

$$\begin{aligned} a &:= 1 - \cos(wT(N+1)) & b &= \sin(wT(N+1)) \\ c &= 1 - \cos(wT) & d &= \sin(wT) \end{aligned} \quad (3.43)$$

Hence, using 3.43, it follows

$$\frac{1 - e^{iwT(N+1)}}{1 - e^{-iwT}} = \frac{(a + ib)}{(c + id)} \quad (3.44)$$

By rearranging the terms, it follows $(a + ib) = (c + id)(p + iq)$ and by multiplying its right hand-side out we get the following system of equations:

$$\begin{aligned} (cp - dq) &= a \\ (dp + cq) &= b \end{aligned} \quad (3.45)$$

After multiplying the first equation of 3.45 by c and the second by d and then adding them together, we get using the law of distributivity new identities for p and q :

$$\begin{aligned} p &= \frac{(ac + bd)}{c^2 + d^2} \\ q &= \frac{(bc + ad)}{c^2 + d^2} \end{aligned} \quad (3.46)$$

Using some trigonometric identities and putting our substitution from 3.43 for a, b, c, d back into the current representation 3.46 of p and q we will get:

$$\begin{aligned} p &= \frac{1}{2} + \frac{1}{2} \left(\frac{\cos(wTN) - \cos(wT(N+1))}{1 - \cos(wT)} \right) \\ q &= \frac{\sin(wT(N+1)) - \sin(wTN) - \sin(wT)}{2(1 - \cos(wT))} \end{aligned} \quad (3.47)$$

Since we have seen, that $\sum_{n=0}^N e^{-iwnT}$ is a complex number and can be written as $(p + iq)$, we now know an explicit expression for p and q . Therefore, the one dimensional inverse Fourier transform of S is equal:

$$\begin{aligned} \mathcal{F}^{-1}\{S\}(w) &= \mathcal{F}^{-1}\{f\}(w) \sum_{n=0}^N e^{-iwnT} \\ &= (p + iq)\mathcal{F}^{-1}\{f\}(w) \end{aligned} \quad (3.48)$$

Now lets consider our actual problem description. Given a patch of a nano-scaled surface snake shed represented as a two dimensional heightfield $h(x, y)$. We once again assume that this provided patch is representing the whole surface S of our geometry by some number of replicas of itself. Therefore, $S(x, y) = \sum_{n=0}^N h(x + nT_1, y + mT_2)$, assuming the given height field has the dimensions T_1 by T_2 . In order to derive an identity for the two dimensional inverse Fourier transformation of S we can similarly proceed like we did to derive equation 3.48.

$$\begin{aligned}
\mathcal{F}^{-1}\{S\}(w_1, w_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{n_2=0}^{N_1} \sum_{n_2=0}^{N_2} h(x_1 + n_1 T_1, x_2 + n_2 T_2) e^{iw(x_1+x_2)} dx_1 dx_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{n_2=0}^{N_1} \sum_{n_2=0}^{N_2} h(y_1, y_2) e^{iw((y_1-n_1 T_1)+(y_2+n_2 T_2))} dx_1 dx_2 \\
&= \sum_{n_2=0}^{N_1} \sum_{n_2=0}^{N_2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(y_1, y_2) e^{iw(y_1+y_2)} e^{-iw(n_1 T_1+n_2 T_2)} dy_1 dy_2 \\
&= \sum_{n_2=0}^{N_1} \sum_{n_2=0}^{N_2} e^{-iw(n_1 T_1+n_2 T_2)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{Box}(y_1, y_2) e^{iw(y_1+y_2)} dy_1 dy_2 \\
&= \left(\sum_{n_2=0}^{N_1} \sum_{n_2=0}^{N_2} e^{-iw(n_1 T_1+n_2 T_2)} \right) \mathcal{F}^{-1}\{h\}(w_1, w_2) \\
&= \left(\sum_{n_2=0}^{N_1} e^{-iwn_1 T_1} \right) \left(\sum_{n_2=0}^{N_2} e^{-iwn_2 T_2} \right) \mathcal{F}^{-1}\{h\}(w_1, w_2) \\
&= (p_1 + iq_1)(p_2 + iq_2) \mathcal{F}^{-1}\{h\}(w_1, w_2) \\
&= ((p_1 p_2 - q_1 q_2) + i(p_1 p_2 + q_1 q_2)) \mathcal{F}^{-1}\{h\}(w_1, w_2) \\
&= (p + iq) \mathcal{F}_{DTFT}^{-1}\{h\}(w_1, w_2)
\end{aligned} \tag{3.49}$$

Where we have defined

$$\begin{aligned}
p &:= (p_1 p_2 - q_1 q_2) \\
q &:= (p_1 p_2 + q_1 q_2)
\end{aligned} \tag{3.50}$$

For the identity of equation 3.49 we made use of Green's integration rule which allowed us to split the double integral to the product of two single integrations. Also, we used the definition of the 2-dimensional inverse Fourier transform of the height field function. We applied a similar substitution like we did in 3.37, but this time twice, once for x_1 and once for x_2 separately. The last step in equation 3.49, substituting with p and q in equation 3.50 will be useful later in the implementation. The insight should be, that the product of two complex numbers is again a complex number. We will have to compute the absolute value of $\mathcal{F}^{-1}\{S\}(w_1, w_2)$ which will then be equal $(p^2 + q^2)^{\frac{1}{2}} |\mathcal{F}^{-1}\{h\}(w_1, w_2)|$

3.10.2 Interpolation

In 3.10.1 we have derived an alternative approach when we are working with a periodic signal instead using the gaussian window approach from sec sec : gaussianwindow. Its main finding 3.49 that we can just integrate over one of its period instead iterating over the whole domain. Nevertheless, this main finding is using the inverse DTFT. Since we are using

We are interested in recovering an original analog signal $x(t)$ from its samples $x[t] =$

Therefore, for a given sequence of real numbers $x[n]$, representing a digital signal, its correspond continuous function is:

$$x(t) = \sum_{n=-\infty}^{\infty} x[n] \text{sinc}\left(\frac{t-nT}{T}\right) \quad (3.51)$$

which has the Fourier transformation $X(f)$ whose non-zero values are confined to the region $|f| \leq \frac{1}{2T} = B$. When $x[n]$ represents time samples at interval T of a continuous function, then the quantity $f_s = \frac{1}{T}$ is known as its sample rate and $\frac{f_s}{2}$ denotes the Nyquist frequency. The sampling Theorem states that when a function has a Bandlimit B less than the Nyquist frequency, then $x(t)$ is a perfect reconstruction of the original function.

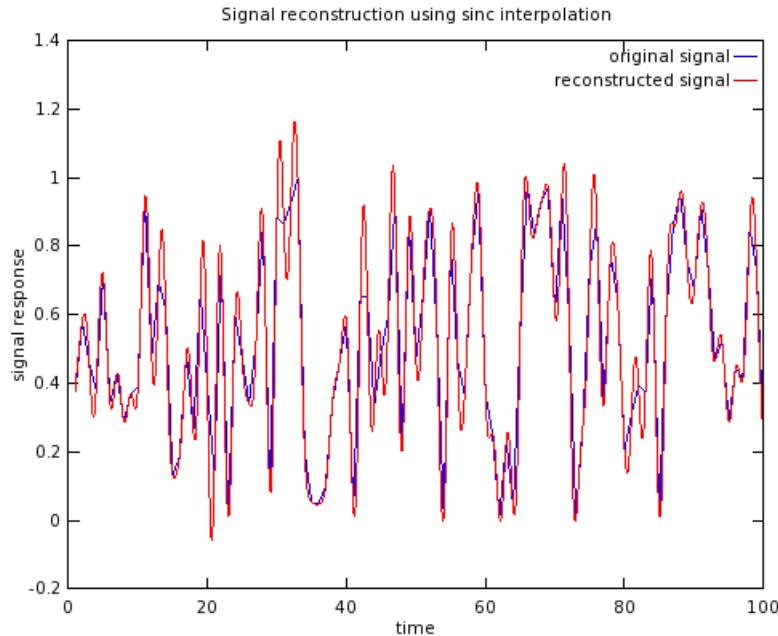


Abbildung 3.4: Comparission between a given random one dimensional input signal $s(t)$ and its sinc interpolation $\hat{s}(t)$. Notice that for the interpolation there were $N = 100$ samples from the original signal provided.

Anhang A

Signal Processing Basics

A signal is a function that conveys information about the behavior or attributes of some phenomenon. In the physical world, any quantity exhibiting variation in time or variation in space (such as an image) is potentially a signal that might provide information on the status of a physical system, or convey a message between observers.

The Fourier Transform is an important image processing tool which is used to decompose an image into its sine and cosine components. The output of the transformation represents the image in the Fourier or frequency domain, while the input image is the spatial domain equivalent. In the Fourier domain image, each point represents a particular frequency contained in the spatial domain image.

A.1 Fourier Transformation

The Fourier-Transform is a mathematical tool which allows to transform a given function or rather a given signal from defined over a time- (or spatial-) domain into its corresponding frequency-domain.

Let f an measurable function over \mathbb{R}^n . Then, the continuous Fourier Transformation(**FT**), denoted as $\mathcal{F}\{f\}$ of f , ignoring all constant factors in the formula, is defined as:

$$\mathcal{F}_{FT}\{f\}(w) = \int_{\mathbb{R}^n} f(x)e^{-iwt}dt \quad (\text{A.1})$$

whereas its inverse transform is defined like the following which allows us to obtain back the original signal:

$$\mathcal{F}_{FT}^{-1}\{f\}(w) = \int_{\mathbb{R}} \mathcal{F}\{w\}e^{iwt}dt \quad (\text{A.2})$$

Usual w is identified by the angular frequency which is equal $w = \frac{2\pi}{T} = 2\pi v_f$. In this connection, T is the period of the resulting spectrum and v_f is its corresponding frequency.

By using Fourier Analysis, which is the approach to approximate any function by sums of simpler trigonometric functions, we gain the so called Discrete Time Fourier Transform (in short **DTFT**). The DTFT operates on a discrete

function. Usually, such an input function is often created by digitally sampling a continuous function. The DTFT itself is operation on a discretized signal on a continuous, periodic frequency domain and looks like the following:

$$\mathcal{F}_{DTFT}\{f\}(w) = \sum_{-\infty}^{\infty} f(x)e^{-iwk} \quad (A.3)$$

Note that the DTFT is not practically suitable for digital signal processing since there a signal can be measured only in a finite number of points. Thus, we can further discretize the frequency domain and will get then the Discrete Fourier Transformation (in short **DFT**) of the input signal:

$$\mathcal{F}_{DFT}\{f\}(w) = \sum_{n=0}^{N-1} f(x)e^{-iw_n k} \quad (A.4)$$

Where the angular frequency w_n is defined like the following $w_n = \frac{2\pi n}{N}$ and N is the number of samples within an equidistant period sampling.

Any continuous function $f(t)$ can be expressed as a series of sines and cosines. This representation is called the Fourier Series (denoted by *FS*) of $f(t)$.

$$f(t) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos(nt) + \sum_{n=1}^{\infty} b_n \sin(nt) \quad (A.5)$$

where

$$\begin{aligned} a_0 &= \int_{-\pi}^{\pi} f(t)dt \\ a_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos(nt) dt \\ b_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin(nt) dt \end{aligned} \quad (A.6)$$

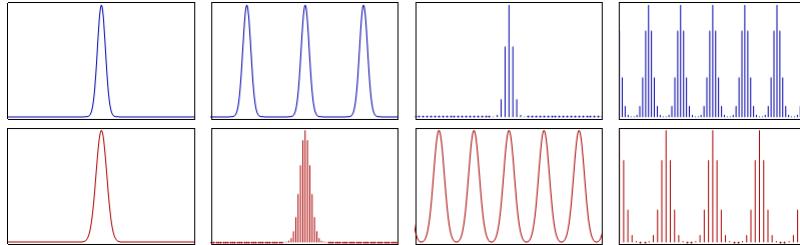


Abbildung A.1: Relationship¹ between the continuous Fourier transform and the discrete Fourier transform: Left column: A continuous function (top) and its Fourier transform *A.1* (bottom). Center-left column: Periodic summation of the original function (top). Fourier transform (bottom) is zero except at discrete points. The inverse transform is a sum of sinusoids called Fourier series *A.5*. Center-right column: Original function is discretized (multiplied by a Dirac comb) (top). Its Fourier transform (bottom) is a periodic summation (DTFT) of the original transform. Right column: The DFT *A.4* (bottom) computes discrete samples of the continuous DTFT *A.3*. The inverse DFT (top) is a periodic summation of the original samples.

Spetial signal $f(t)$ is	Operator	Transformed frequency signal $\hat{f}(\omega)$ is
continuous and periodic in t	FS A.5	only discrete in ω
only continuous in t	FT A.1	only continuous in ω
only discrete in t	DTFT A.3	continuous and periodic in ω
discrete and periodic in t	DFT A.4	discrete and periodic in ω

Tabelle A.1: Fourier operator to apply for a given spatial input signal and the properties of its resulting output signal in frequency space

A.2 Convolution

The convolution $f * g$ of two functions $f, g: \mathbb{R}^n \rightarrow \mathbb{C}$ is defined as:

$$(f * g)(t) = \int_{\mathbb{R}^n} f(t)g(t - x)dx \quad (\text{A.7})$$

Note that the Fourier transform of the convolution of two functions is the product of their Fourier transforms. This is equivalent to the fact that Convolution in spatial domain is equivalent to multiplication in frequency domain. Therefore, the inverse Fourier transform of the product of two Fourier transforms is the convolution of the two inverse Fourier transforms. Last an illustration of the relationships between the previous presented Fourier transformations and different given input signals. First an concrete example shown in Figure A.1. Table A.1 tells what Fourier transformation operator has to be applied to which kind of input signal and what properties its resulting Fourier transform will have.

A.3 Taylor Series

Taylor series is a representation of a function as an infinite sum of terms that are calculated from the values of the function's derivatives at a single point.

The Taylor series \mathcal{T} of a real or complex-valued function $f(x)$ that is infinitely differentiable at a real or complex number a is the power series:

$$\mathcal{T}(f; a)(x) = \sum_{n=0}^{\infty} \frac{f^n(a)}{n!} (x - a)^n \quad (\text{A.8})$$

¹image of illustration has been taken from wikipedia

Anhang B

Summary of Stam's Derivations

In his paper about Diffraction Shader, J. Stam derives a BRDF which is modeling the effect of diffraction for various analytical anisotropic reflexion models relying on the so called scalar wave theory of diffraction for which a wave is assumed to be a complex valued scalar. It's noteworthy, that Stam's BRDF formulation does not take into account the polarization of the light. Fortunately, light sources like sunlight and light bulbs are unpolarized.

A further assumption in Stam's Paper is, the emanated waves from the source are stationary, which implies the wave is a superposition of independent monochromatic waves. This implies that each wave is associated to a definite wavelength lambda. However, sunlight once again fulfills this fact.

In our simulations we will always assume we have given a directional light source, i.e. sunlight. Hence, Stam's model can be used for our derivations.

For his derivations Stam uses the Kirchhoff integral¹, which is relating the reflected field to the incoming field. This equation is a formalization of Huygen's well-known principle that states that if one knows the wavefront at a given moment, the wave at a later time can be deduced by considering each point on the first wave as the source of a new disturbance. Mathematically speaking, once the field $\psi_1 = e^{ik\mathbf{x} \cdot \mathbf{s}}$ on the surface is known, the field ψ_2 everywhere else away from the surface can be computed. More precisely, we want to compute the wave ψ_2 equal to the reflection of an incoming planar monochromatic wave $\psi_1 = e^{ik\omega_i * x}$ traveling in the direction ω_i from a surface S to the light source. Formally, this can be written as:

$$\psi_2(\omega_i, \omega_r) = \frac{ik e^{iKR}}{4\pi R} (F(-\omega_i - \omega_r) - (-\omega_i + \omega_r)) \cdot I_1(\omega_i, \omega_r) \quad (\text{B.1})$$

with

$$I_1(\omega_i, \omega_r) = \int_S \hat{\mathbf{n}} e^{ik(-\omega_i - \omega_r) \cdot \mathbf{s}} d\mathbf{s} \quad (\text{B.2})$$

¹See http://en.wikipedia.org/wiki/Kirchhoff_integral_theorem for further information.

In applied optics, when dealing with scattered waves, one does use differential scattering cross-section rather than defining a BRDF which has the following identity:

$$\sigma^0 = 4\pi \lim_{R \rightarrow \infty} R^2 \frac{\langle |\psi_2|^2 \rangle}{\langle |\psi_1|^2 \rangle} \quad (\text{B.3})$$

where R is the distance from the center of the patch to the receiving point x_p , \hat{n} is the normal of the surface at s and the vectors:

The relationship between the BRDF and the scattering cross section can be shown to be equal to

$$\text{BRDF} = \frac{1}{4\pi A} \frac{1}{\cos(\theta_i)\cos(\theta_r)} \frac{\sigma^0}{\langle |\psi_1|^2 \rangle} \quad (\text{B.4})$$

where θ_i and θ_r are the angles of incident and reflected directions on the surface with the surface normal n . See 2.13.

The components of vector resulting by the difference between these direction vectors: In order to simplify the calculations involved in his vectorized integral equations, Stam considers the components of vector

$$(u, v, w) = -\omega_i - \omega_r \quad (\text{B.5})$$

explicitly and introduces the equation:

$$I(ku, kv) = \int_S \hat{n} e^{ik(u, v, w) \cdot s} ds \quad (\text{B.6})$$

which is a first simplification of B.2. Note that the scalar w is the third component of 2.5 and can be written as $w = -(\cos(\theta_i) + \cos(\theta_r))$ using spherical coordinates. The scalar $k = \frac{2\pi}{\lambda}$ represent the wavenumber.

During his derivations, Stam provides a analytical representation for the Kirchhoff integral assuming that each surface point $s(x, y)$ can be parameterized by $(x, y, h(x, y))$ where h is the height at the position (x, y) on the given (x, y) surface plane. Using the tangent plane approximation for the parameterized surface and plugging it into B.6 he will end up with:

$$\mathbf{I}(ku, kv) = \int \int (-h_x(x, y), -h_y(x, y), 1) e^{ikwh(x, y)} e^{ik(ux+vy)} dx dy \quad (\text{B.7})$$

For further simplification Stam formulates auxillary function which depends on the provided height field:

$$p(x, y) = e^{ikwh(x, y)} \quad (\text{B.8})$$

which will allow him to further simplify his equation B.7 to:

$$\mathbf{I}(ku, kv) = \int \int \frac{1}{ikw} (-p_x, -p_y, ikwp) dx dy \quad (\text{B.9})$$

where he used that $(-h_x(x, y), -h_y(x, y), 1) e^{ikwh(x, y)}$ is equal to $\frac{(-p_x, -p_y, ikwp)}{ikw}$ using the definition of the partial derivatives applied to the function 2.7.

Let $P(x, y)$ denote the Fourier Transform (FT) of $p(x, y)$. Then, the differentiation with respect to x respectively to y in the Fourier domain is equivalent to a multiplication of the Fourier transform by $-iku$ or $-ikv$ respectively. This leads him to the following simplification for *B.7*:

$$\mathbf{I}(ku, kv) = \frac{1}{w} P(ku, kv) \cdot (u, v, w) \quad (\text{B.10})$$

Let us consider the term $g = (F(-\omega_i - \omega_r) - (-\omega_i + \omega_r))$, which is a scalar factor of *B.1*. The dot product with g and $(-\omega_i - \omega_r)$ is equal $2F(1 + \omega_i \cdot \omega_r)$. Putting this finding and the identity *B.10* into *B.1* he will end up with:

$$\psi_2(\omega_i, \omega_r) = \frac{ike^{iKR}}{4\pi R} \frac{2F(1 + \omega_i \cdot \omega_r)}{w} P(ku, kv) \quad (\text{B.11})$$

By using the identity *B.4*, this will lead us to his main finding:

$$BRDF_\lambda(\omega_i, \omega_r) = \frac{k^2 F^2 G}{4\pi^2 A w^2} \langle |P(ku, kv)|^2 \rangle \quad (\text{B.12})$$

where G is the so called geometry term which is equal:

$$G = \frac{(1 + \omega_i \cdot \omega_r)^2}{\cos(\theta_i) \cos(\theta_r)} \quad (\text{B.13})$$

Anhang C

Appendix

C.1 Schlick's approximation

The Fresnel's equations describe the reflection and transmission of electromagnetic waves at an interface. That is, they give the reflection and transmission coefficients for waves parallel and perpendicular to the plane of incidence. Schlick's approximation is a formula for approximating the contribution of the Fresnel term where the specular reflection coefficient R can be approximated by:

$$R(\theta) = R_0 + (1 - R_0)(1 - \cos \theta)^5 \quad (\text{C.1})$$

and

$$R_0 = \left(\frac{n_1 - n_2}{n_1 + n_2} \right)^2$$

where θ is the angle between the viewing direction and the half-angle direction, which is halfway between the incident light direction and the viewing direction, hence $\cos \theta = (H \cdot V)$. And n_1, n_2 are the indices of refraction of the two medias at the interface and R_0 is the reflection coefficient for light incoming parallel to the normal (i.e., the value of the Fresnel term when $\theta = 0$ or minimal reflection). In computer graphics, one of the interfaces is usually air, meaning that n_1 very well can be approximated as 1.

C.2 Spherical Coordinates

$$\forall \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 : \exists r \in [0, \infty) \exists \phi \in [0, 2\pi] \exists \theta \in [0, \pi] \text{ s.t.}$$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r \sin(\theta) \cos(\phi) \\ r \sin(\theta) \sin(\phi) \\ r \cos(\theta) \end{pmatrix}$$

C.3 Tangent Space

The concept of tangentspace-transformation of tangent space is used in order to convert a point between world and tangent space. GLSL fragment shaders require normals and other vertex primitives declared at each pixel point, which mean that we have one normal vector at each texel and the normal vector axis will vary for every texel.

Think of it as a bumpy surface defined on a flat plane. If those normals were declared in the world space coordinate system, we would have to rotate these normals every time the model is rotated, even when just for a small amount. Since the lights, cameras and other objects are usually defined in world space coordinate system, and therefore, when they are involved in a calculation within the fragment shader, we would have to rotate them as well for every pixel. This would involve almost countless many object to world matrix transformations needed to take place at the pixel level. Therefore, instead doing so, we transform all vertex primitives into tangent space within the vertex shader.

To make this point clear an example: Even we would rotate the cube in figure C.1, the tangent space axis will remain aligned with respect to the face. Which practically speaking, will save us from performing many space transformations applied pixel-wise within the fragment shader and instead allows us to perform us the tangenspace transformation of every involved vertex primitive in the vertex-shader.

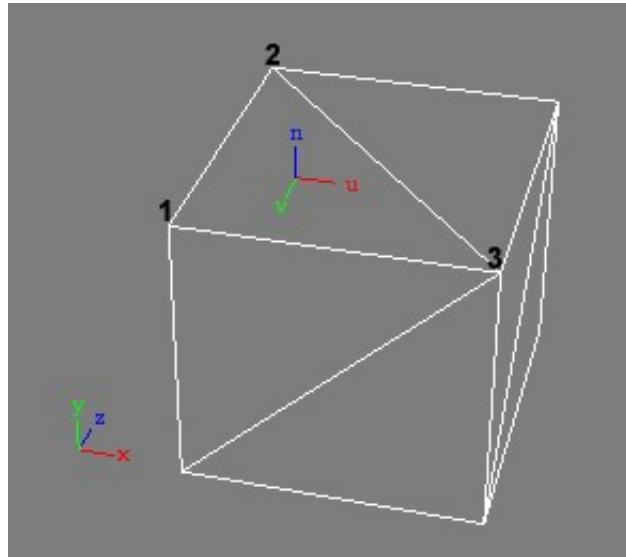


Abbildung C.1: Cube in world space (x, y, z) showing the tangent space (u, v, n) of its face (2, 1, 3)

Tabellenverzeichnis

A.1 Fourier operator to apply for a given spatial input signal and the properties of its resulting output signal in frequency space	44
---	----

Abbildungsverzeichnis

1.1	example biological color production	1
1.2	Structural color examples	2
1.3	Xenopeltis AFM image	3
2.1	Irradiance is the summed up radiance over all directions	8
2.2	BRDF Model	9
2.3	visiblelightspectrum	10
2.4	humanayeschematic	11
2.5	Plots of our color matching functions we used for rendering	12
2.6	sinewave	14
2.7	interference	15
2.8	Two mathematical sine waves which are perfectly coherent which means that their phase difference is constant for every point in time.	16
2.9	Huygen's Principle	18
2.10	Diffracted Wave	19
2.11	diffractiondiemsnion	20
2.12	Idea behind Stam's approach	21
2.13	Stam's geometrical setup	22
2.14	Comparing Stam's apporach: Gratings	23
2.15	Comparing Stam's apporach: Good Example	24
2.16	Comparing Stam's apporach: Bad Example	24
3.1	Problem Statement	26
3.2	Problem Statement: Output	27
3.3	FT by DTFT	29
3.4	Comparission between a given random one dimensional input signal $s(t)$ and its sinc interpolation $\hat{s}(t)$. Notice that for the interpolation there were $N = 100$ samples from the original signal provided.	41
C.1	Cube in world space (x, y, z) showing the tangen space (u, v, n) of its face (2,1,3)	49

List of Algorithms

Literaturverzeichnis

- [Bar07] BARTSCH, Hans-Jochen: *Taschenbuch Mathematischer Formeln*. 21th edition. HASNER, 2007. – ISBN 978–3–8348–1232–2
- [CT12] CUYPERS T., et a.: Reflectance Model for Diffraction. In: *ACM Trans. Graph.* 31, 5 (2012), September
- [DSD14] D. S. DHILLON, et a.: Interactive Diffraction from Biological Nanostructures. In: *EUROGRAPHICS 2014/ M. Paulin and C. Dachsba-cher* (2014), January
- [For11] FORSTER, Otto: *Analysis 3*. 6th edition. VIEWEG+TEUBNER, 2011. – ISBN 978–3–8348–1232–2
- [I.N14] I.NEWTON: *Opticks, reprinted*. CreateSpace Independent Publishing Platform, 2014. – ISBN 978–1499151312
- [JG04] JUAN GUARDADO, NVIDIA: Simulating Diffraction. In: *GPU Gems* (2004). <https://developer.nvidia.com/content/gpu-gems-chapter-8-simulating-diffraction>
- [MT10] MATIN T.R., et a.: Correlating Nanostructures with Function: Structural Colors on the Wings of a Malaysian Bee. (2010), August
- [PAT09] PAUL A. TIPLER, Gene M.: *Physik für Wissenschaftler und Ingenieure*. 6th edition. Spektrum Verlag, 2009. – ISBN 978–3–8274–1945–3
- [PS09] P. SHIRLEY, S. M.: *Fundamentals of Computer Graphics*. 3rd edition. A K Peters, Ltd, 2009. – ISBN 978–1–56881–469–8
- [R.H12] R.HOOKE: *Micrographia, reprinted*. CreateSpace Independent Publishing Platform, 2012. – ISBN 978–1470079031
- [RW11] R. WRIGHT, et a.: *OpenGL SuperBible*. 5th edition. Addison-Wesley, 2011. – ISBN 978–0–32–171261–5
- [Sta99] STAM, J.: Diffraction Shaders. In: *SIGGRPAH 99 Conference Procee-dings* (1999), August

Erklärung

gemäss Art. 28 Abs. 2 RSL 05

Name/Vorname:

Matrikelnummer:

Studiengang:

Bachelor

Master

Dissertation

Titel der Arbeit:

.....
.....

LeiterIn der Arbeit:

.....

Ich erkläre hiermit, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls der Senat gemäss Artikel 36 Absatz 1 Buchstabe o des Gesetztes vom 5. September 1996 über die Universität zum Entzug des auf Grund dieser Arbeit verliehenen Titels berechtigt ist.

.....
Ort/Datum

.....
Unterschrift