

Diffraction Shader

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	2
1.3	Previous work	2
1.4	Overview	3
2	Theoretical Background	4
2.1	Basics in Modeling Light in Computer Graphics	4
2.1.1	Radiometry	4
2.1.2	Spectral Energy	4
2.1.3	Spectral Power	5
2.1.4	Spectral Irradiance	5
2.1.5	Spectral Radiance	6
2.1.6	BRDF	7
2.1.7	Wavespectrum and Colors	7
2.1.8	Colorspace	9
2.1.9	Spectral Rendering	10
2.2	Wave Theory for Light and Diffraction	10
2.2.1	Basics in Wave Theory	10
2.2.2	Wave Coherence	12
2.2.3	Hyugens Principle	12
2.2.4	Diffraction of Waves	12
2.3	Stam's BRDF formulation	12
3	Derivations	15
3.1	Adaption of Stam's BRDF	15
3.1.1	BRDF formulation	15
3.1.2	Relative BRDF	16
3.1.3	Taylor approximation for BRDF	18
3.1.4	Sampling: Gaussian Window	22
3.1.5	Final Expression	22
3.2	Alternative Approach	23
3.2.1	PQ factors	23
3.2.2	Interpolation	26

A Appendix	28
A.1 Signal Processing Basics	28
A.1.1 Fourier Transformation	28
A.1.2 Convolution	30
A.1.3 Taylor Series	30
B Appendix	32
B.1 Schlick's approximation	32
B.2 Spherical Coordinates	32
B.3 Tangent Space	33
List of Tables	34
List of Figures	34
List of Algorithms	35
Bibliography	36

Kapitel 1

Introduction

1.1 Motivation

In Nature, coloring mostly comes from the inherent colors of materials but sometimes colorization has a pure physical origin such as the effect diffraction or interference of light. Both phenomenon are causing the so called structural coloration, which is the production of color through the interaction of visible light with micrioscopically structured surfaces. Color production is due to wave interference with quasiperiodic structures whose periodicity leads to interaction with visible light. Therefore we perceive color when the different wavelengths composing white light are selectively interfered with by matter (absorbed, reflected, refracted, scattered, or diffracted) on their way to our eyes, or when a non-white distribution of light has been emitted. In animals, such as feathers of birds and the scales of butterflies, interference is created by a range of photonic mechanisms, including diffraction grating, selective mirrors, photonic crystals. The connection between microscopic structures and coloration has been observed by Robert Hooke in the early seventeenth century. The discovery of the wave nature of light led to the conclusion that the cause for the coloration lies in wave interference.

In the field of computer graphics, many researchers have been attempted rendering of structural colors by formulating a the bidirectional reflectance distribution function (BRDF) for this purpose. But most of the techniques so far, however, are either too slow for interactive rendering or rely on simplifying assumption, like modeling light as rays, to achieve real-time performance, which are not able capturing the essence of diffraction at all.



(a) Elaphe Guttata Snake



(b) Xenopeltis Snake

Abbildung 1.1: Effect of diffraction on snake sheds for different species

1.2 Goals

The purpose of this thesis is to simulate realistically by rendering structural colors caused by the effect of diffraction on different biological structures in realtime. We focus on structural colors generated by diffraction gratings, in particular our approach applies to surfaces with quasiperiodic structures at the nanometer scale that can be represented as heightfields. such structures are found on the sehs of snakes, wings of butterflies or the bodies of various insects. we restrict ourself and focus on different snake skins sheds which are acquired nanoscaled heightfields using atomic force microscopy.

In oder to achieve our rendering purpose we will rely J. Stam's formulation of a BRDF which basically describes the effect of diffraction on a given surface assuming one knows the hightfield of this surface and will further extend this. Appart from Stam's approach, which models the heightfield as a probabilistic superposition of bumps and proceeds to derive an analytical expression for the BRDF, our BRDF representation takes the heightfield from explicit measurement. I.E. in our case, those heightfields are small patches of the microstructured surfaces (in nano-scale) taken by AFM of snake skin patches provided by our collaborators in Geneva.. So this approach is closer to real truth, since we use measured surfaces instead of statistical surface profile.

Therefore, this work can be considered as an extension of J. Stam's derivations for the case one is provided by a explicit height field on a quasiperiodic structure.

Real time performance is achieved with a representation of the formula as a power series over a variable related to the viewing and lighting directions. Values closely related to the coefficients in that power series are precomputed.

The contribution is that this approach is more broadly applicable than the previous work. Although the previously published formula theoretically has this much flexibility already, there is a novel contribution in demonstrating how such generality can be leveraged in practical implementation

1.3 Previous work

stam, hooke, see our paper, see stams paper, see own research.

Robert Hooke = observed connection between microscopic structures and colorisation wave nature of light led to conclusion that the cause for the colouration lies in wave interference.

previous

In computer graphics literature, Stam was the first to develop reflection models based on wave optics called diffraction shaders, that can produce colorful diffraction effects. His approach is based on a far field approximation of the Kirchhoff integral. He shows that for surfaces represented as nanoscale heightfields it is possible to derive their BRDF as the Fourier transformation of a function of the heightfield. Nevertheless, this formulation is not immediately useful for efficient rendering of measured complex nanostructures since this would require the on-the-fly evaluation of and integration over Fourier transforms of the heightfield that depend on the light and viewing geometry. In his derivations, Stam models heightfields as probabilistic superpositions of bumps forming periodic like structures. This provides him an analytical identity for this class of heightfields. However, biological nanostructures are way more complex and do not lend themselves to this simplified statistical model.

follow ups

1.4 Overview

The remainder of this thesis is organized as follows: due to the fact that this thesis has a rather advanced mathematical complexity the first part of chapter 2 introduces some important definitions which are required in order to be able to follow the derivations in the last third of chapter 2. Before starting the derivations, a brief summary of J. Stam's Paper about diffraction shaders is provided since this whole thesis is based on his BRDF representation. Our derivations itself are listed step-wise, whereas there is a final representation provided by the end of chapter 2. Chapter 3 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 4 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 5 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. Chapter 6 contains the conclusion of this thesis which starts by a review briefly discussing what has been achieved in this thesis and the drawbacks. There are also 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 progresses. 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 distribution ¹.

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 $\Phi_\lambda = \frac{Q}{\lambda}$, 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:

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

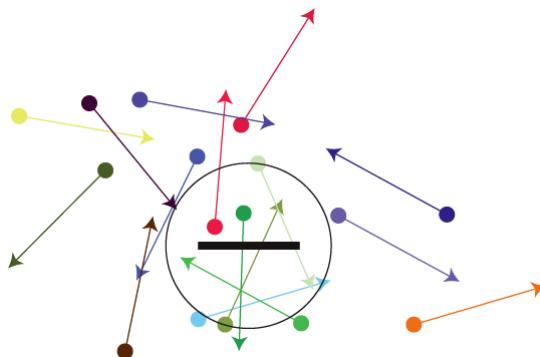
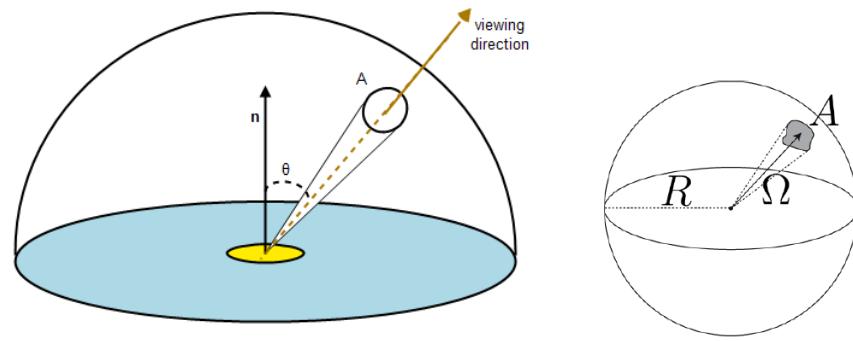


Abbildung 2.1: Irradiance is the summed up radiance over all directions

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

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) Radiance 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

s

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:

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

where L is the observed spectral radiance in the unit energy per unit area per solid angle, which is $Wm^{-2}sr^{-1}$ 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.

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

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.

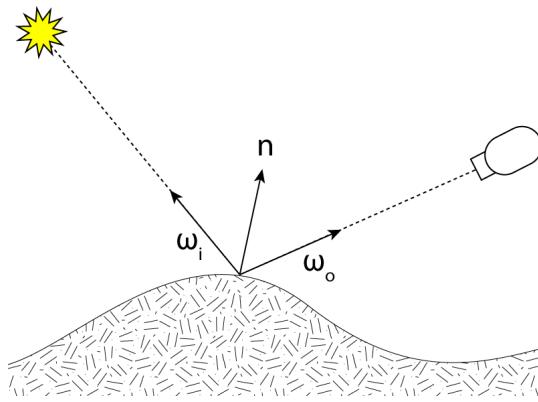


Abbildung 2.2: Illustration³ of the BRDF model, where ω_i is pointing to the light source and the existing direction is denoted by ω_r . Both direction unit direction vectors defined w.r.t to a surface normal \mathbf{n} for every point on the surface.

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*⁴ 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*

³image of illustration has been taken from wikipedia

⁴mentioned in Computer Graphics Fundamentals Book from the year 2000

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.

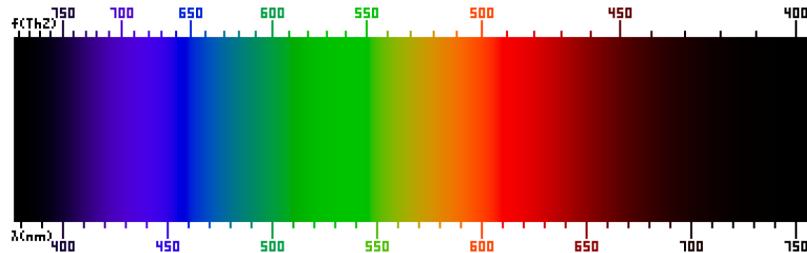


Abbildung 2.3: Frequency (top) and wavelength (bottom) of colors of the visible spectrum⁵.

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.

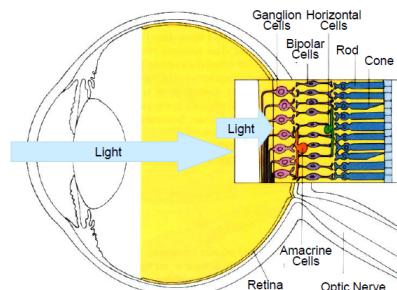


Abbildung 2.4: Schematic⁶ of photoreceptor cells, cones and rods, in human eye

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.

⁵Similar figure like used in computer graphics class 2012 in chapter colors

⁶image of illustration has been taken from wikipedia

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 numbers⁷, 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 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.

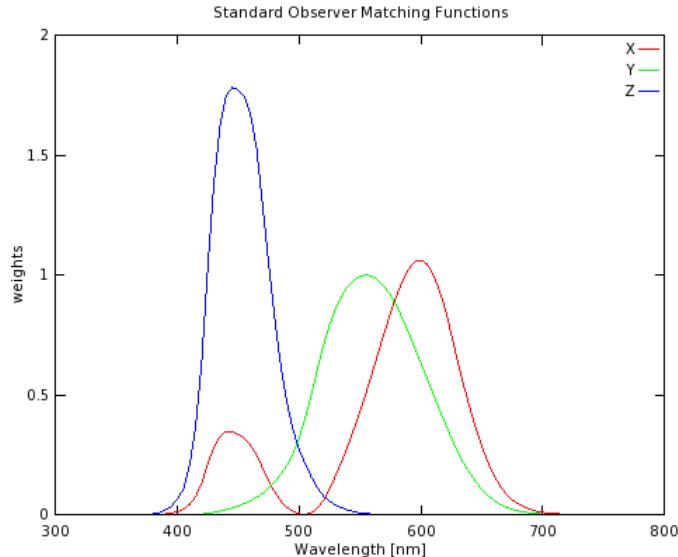


Abbildung 2.5: Plots of our color matching functions we used for rendering

⁷note 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.

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 themself 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:

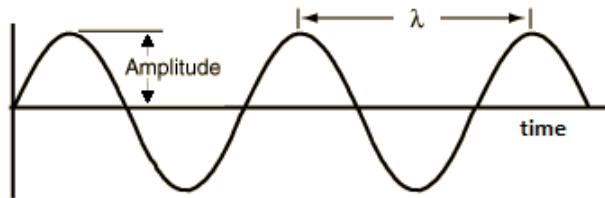


Abbildung 2.6: Simplified, one dimensionaly real valued wave function⁸, giving an idea about the properties of a wave.

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: There are two possible interpretations. First, it is a measure of the height from the equilibrium point to the highest point of a crest on the wave. For light waves it is a relative measure of intensity or brightness to other light waves of the same wavelength. And secondly, it can be interpreted as a measure how much energy a wave carries where the greater the absolute amplitude value, the bigger the amount of energy being carried.

Frequency: Is a measure of the number of waves passing through a point in the propagation medium during a certain time.

Phase: describes either the offset of a wave's initial position 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.

A geometrical property of a wave is the wavefront. This is either a surface or line along the path of wave motion on which the disturbance at every point has the same phase. Note that for electromagnetic waves, the phase is the definition of position of a point in time on a wavefront cycle whereat a complete cycle is defined as being 360 degrees.

⁸Modified, schematically sine-wave

2.2.2 Wave Coherence

2.2.3 Huygens Principle

2.2.4 Diffraction of Waves

2.3 Stam's BRDF formulation

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 integral (ADD REF TO WIKI), 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 $\psi_1 = e^{ik\omega_i \cdot \mathbf{x}}$ traveling in the direction ω_i from a surface S to the light source. Formally, this can be written as:

$$\psi_2(\omega_i, \omega_r) = \frac{ik e^{iKR}}{4\pi R} (F(-\omega_i - \omega_r) - (-\omega_i + \omega_r)) \cdot I_1(\omega_i, \omega_r) \quad (2.5)$$

with

$$I_1(\omega_i, \omega_r) = \int_S \hat{\mathbf{n}} e^{ik(-\omega_i - \omega_r) \cdot \mathbf{s}} d\mathbf{s} \quad (2.6)$$

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:

$$\sigma^0 = 4\pi \lim_{R \rightarrow \infty} R^2 \frac{\langle |\psi_2|^2 \rangle}{\langle |\psi_1|^2 \rangle} \quad (2.7)$$

where R is the distance from the center of the patch to the receiving point x_p , $\hat{\mathbf{n}}$ 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

$$BRDF = \frac{1}{4\pi} \frac{1}{A} \frac{\sigma^0}{\cos(\theta_i)\cos(\theta_r)} \quad (2.8)$$

where θ_i and θ_r are the angles of incident and reflected directions on the surface with the surface normal n . See 2.7.

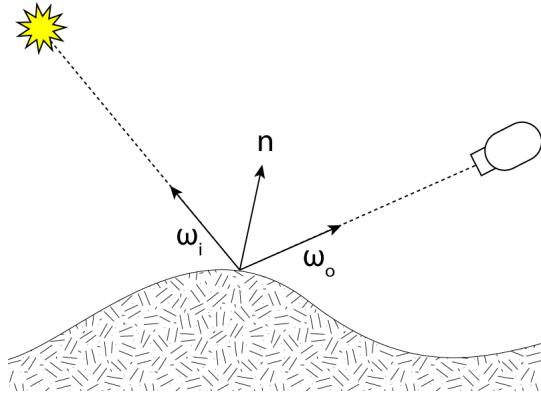


Abbildung 2.7: ω_i points toward the light source, ω_r points toward the camera, n is the surface normal

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

$$(u, v, w) = -\omega_i - \omega_r \quad (2.9)$$

explicitly and introduces the equation:

$$I(ku, kv) = \int \hat{n} e^{ik(u, v, w) \cdot \mathbf{s} d\mathbf{s}} \quad (2.10)$$

which is a first simplification of 2.6. Note that the scalar w is the third component of 2.9 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 2.10 he will end up with:

$$\mathbf{I}(ku, kv) = \int \int (-h_x(x, y), -h_y(x, y), 1) e^{ikwh(x, y)} e^{ik(ux+vy)} dx dy \quad (2.11)$$

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

$$p(x, y) = e^{iwkh(x, y)} \quad (2.12)$$

which will allow him to further simplify his equation 2.11 to:

$$\mathbf{I}(ku, kv) = \int \int \frac{1}{ikw} (-p_x, -p_y, ikwp) dx dy \quad (2.13)$$

where he used that $(-h_x(x, y), -h_y(x, y), 1)e^{kwh(x, y)}$ is equal to $\frac{(-p_x, -p_y, ikwp)}{ikw}$ using the definition of the partial derivatives applied to the function 2.12.

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 2.11:

$$\mathbf{I}(ku, kv) = \frac{1}{w} P(ku, kv) \cdot (u, v, w) \quad (2.14)$$

Let us consider the term $g = (F(-\omega_i - \omega_r) - (-\omega_i + \omega_r))$, which is a scalar factor of 2.5. 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 2.14 into 2.5 he will end up with:

$$\psi_2(\omega_i, \omega_r) = \frac{ike^{iKR}}{4\pi R} \frac{2F(1 + \omega_i \cdot \omega_r)}{w} P(ku, kv) \quad (2.15)$$

By using the identity 2.8, this will lead us to his main finding:

$$BRDF_\lambda(\omega_i, \omega_r) = \frac{k^2 F^2 G}{4\pi^2 Aw^2} \langle |P(ku, kv)|^2 \rangle \quad (2.16)$$

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

$$G = \frac{(1 + \omega_i \cdot \omega_r)^2}{\cos(\theta_i)\cos(\theta_r)} \quad (2.17)$$

Kapitel 3

Derivations

3.1 Adaption of Stam's BRDF

3.1.1 BRDF formulation

Lets assume we have given an incoming light source with solid angle ω_i , 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:

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

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 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 $\omega = \omega_i$.

We are going to use Stam's main derivation (2.16) for the $BRDF(\omega_i, \omega_r)$ in 3.3 by applying the fact that the wavenumber is equal $k = \frac{2\pi}{\lambda}$:

$$\begin{aligned} BRDF(\omega_i, \omega_r) &= \frac{k^2 F^2 G}{4\pi^2 A w^2} \langle |P(ku, kv)|^2 \rangle \\ &= \frac{k^2 F^2 (1 + \omega_i \cdot \omega_r)^2}{\cos(\theta_i) \cos(\theta_r) 4\pi^2 A w^2} \langle |P(ku, kv)|^2 \rangle \\ &= \frac{4\pi^2 F^2 (1 + \omega_i \cdot \omega_r)^2}{\cos(\theta_i) \cos(\theta_r) 4\pi^2 A \lambda^2 w^2} \langle |P(ku, kv)|^2 \rangle \\ &= \frac{F(\omega_i, \omega_r)^2 (1 + \omega_i \cdot \omega_r)^2}{\cos(\theta_i) \cos(\theta_r) A \lambda^2 w^2} \langle |P(ku, kv)|^2 \rangle \end{aligned} \quad (3.4)$$

Going back to the definition 2.9 of $(u, v, w) = -\omega_i - \omega_r$ and using spherical coordinates B.2, we get for w the following identity

$$\begin{aligned} w &= -\omega_i - \omega_r \\ &= -(\omega_i + \omega_r) \\ &= -(\cos(\theta_i) + \cos(\theta_r)) \end{aligned} \quad (3.5)$$

and therefore w^2 is equal $(\cos(\theta_i) + \cos(\theta_r))^2$. This new fact will allow us to get even further:

$$\begin{aligned} L_\lambda(\omega_r) &= \frac{F(\omega_i, \omega_r)^2 (1 + \omega_i \cdot \omega_r)^2}{A \lambda^2 \cos(\theta_i) \cos(\theta_r) (\cos(\theta_i) + \cos(\theta_r))^2} \left\langle \left| P_{cont} \left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda} \right) \right|^2 \right\rangle \cos(\theta_i) I(\lambda) \\ &= I(\lambda) \frac{F(\omega_i, \omega_r)^2 (1 + \omega_i \cdot \omega_r)^2}{\lambda^2 A (\cos(\theta_i) + \cos(\theta_r))^2 \cos(\theta_r)} \left\langle \left| P_{cont} \left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda} \right) \right|^2 \right\rangle \\ &= I(\lambda) \frac{F(\omega_i, \omega_r)^2 (1 + \omega_i \cdot \omega_r)^2}{\lambda^2 A (\cos(\theta_i) + \cos(\theta_r))^2 \cos(\theta_r)} \left\langle \left| T_0^2 P_{dtft} \left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda} \right) \right|^2 \right\rangle \end{aligned} \quad (3.6)$$

Where P_{cont} denotes the continuous inverse Fourier-Transform A.2 for the Taylor-Series A.8 of our height field representing the nano-scaled surface structure, i.e. $P(k, l) = \mathcal{F}^{-1}\{p\}(k, l)$ and P_{dtft} is the inverse Discrete Time Fourier Transform A.3 of $p(x, y) = e^{ikwh(x, y)}$. Furthermore T_0 the sampling distance for the discretization of $p(x, y)$ assuming equal and uniform sampling in both dimensions x and y .

3.1.2 Relative BRDF

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 BRDF 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 3.6 by simply taking the fraction between $L_\lambda(\omega_r)$ and $L_\lambda^{spec}(\omega_0)$ which is denoted by:

$$\rho_\lambda(\omega_i, \omega_r) = \frac{L_\lambda(\omega_r)}{L_\lambda^{spec}(\omega_0)} \quad (3.7)$$

But first, let us derive the following expression:

$$\begin{aligned} L_\lambda^{spec}(\omega_0) &= I(\lambda) \frac{F(\omega_0, \omega_0)^2 (1 + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix})^2}{\lambda^2 A (\cos(0) + \cos(0))^2 \cos(0)} \langle \left| T_0^2 P_{dtft}(0, 0) \right|^2 \rangle \\ &= I(\lambda) \frac{F(\omega_0, \omega_0)^2 (1+1)^2}{\lambda^2 A (1+1)^2 1} \left| T_0^2 N_{sample} \right|^2 \\ &= I(\lambda) \frac{F(\omega_0, \omega_0)^2}{\lambda^2 A} \left| T_0^2 N_{sample} \right|^2 \end{aligned} \quad (3.8)$$

Where $N_{samples}$ is the number of samples of the DTFT A.3. Thus, we can plug our last derived expression 3.8 into the definition for the relative reflectance radiance 3.7 in the direction w_r and will get:

$$\begin{aligned} \rho_\lambda(\omega_i, \omega_r) &= \frac{L_\lambda(\omega_r)}{L_\lambda^{spec}(\omega_0)} \\ &= \frac{I(\lambda) \frac{F(\omega_i, \omega_r)^2 (1+\omega_i \cdot \omega_r)^2}{\lambda^2 A (\cos(\theta_i) + \cos(\theta_r))^2 \cos(\theta_r)} \langle \left| T_0^2 P_{dtft}(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda}) \right|^2 \rangle}{I(\lambda) \frac{F(\omega_0, \omega_0)^2}{\lambda^2 A} \left| T_0^2 N_{sample} \right|^2} \\ &= \frac{F^2(\omega_i, \omega_r) (1 + \omega_i \cdot \omega_r)^2}{F^2(\omega_0, \omega_0) (\cos(\theta_i) + \cos(\theta_r))^2 \cos(\theta_r)} \langle \left| \frac{P_{dtft}(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda})}{N_{samples}} \right|^2 \rangle \end{aligned} \quad (3.9)$$

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

$$C(\omega_i, \omega_r) = \frac{F^2(\omega_i, \omega_r) (1 + \omega_i \cdot \omega_r)^2}{F^2(\omega_0, \omega_0) (\cos(\theta_i) + \cos(\theta_r))^2 \cos(\theta_r) N_{samples}^2} \quad (3.10)$$

Using this substitute, we will end up with the following expression for the relative reflectance radiance from equation 3.9:

$$\rho_\lambda(\omega_i, \omega_r) = C(\omega_i, \omega_r) \langle \left| P_{dtft}(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda}) \right|^2 \rangle \quad (3.11)$$

Using the previous definition for the relative reflectance radiance 3.7:

$$\rho_\lambda(\omega_i, \omega_r) = \frac{L_\lambda(\omega_r)}{L_\lambda^{spec}(\omega_0)} \quad (3.12)$$

Which we can rearrange to the expression:

$$L_\lambda(\omega_r) = \rho_\lambda(\omega_i, \omega_r) L_\lambda^{spec}(\omega_0) \quad (3.13)$$

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:

$$\begin{aligned} X &= \int_\lambda L_\lambda(\omega_r) \bar{x}(\lambda) d\lambda \\ Y &= \int_\lambda L_\lambda(\omega_r) \bar{y}(\lambda) d\lambda \\ Z &= \int_\lambda L_\lambda(\omega_r) \bar{z}(\lambda) d\lambda \end{aligned} \quad (3.14)$$

where \bar{x} , \bar{y} , \bar{z} are the color matching functions. Using our last finding 3.13 for $L_\lambda(\omega_r)$ with the definition for the tristimulus values 3.14, 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:

$$\begin{aligned} X &= \int_\lambda L_\lambda(\omega_r) \bar{x}(\lambda) d\lambda \\ &= \int_\lambda \rho_\lambda(\omega_i, \omega_r) L_\lambda^{spec}(\omega_0) \bar{x}(\lambda) d\lambda \\ &= \int_\lambda \rho_\lambda(\omega_i, \omega_r) S(\lambda) \bar{x}(\lambda) d\lambda \\ &= \int_\lambda C(\omega_i, \omega_r) \left\langle \left| P_{dtft} \left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda} \right) \right|^2 \right\rangle S(\lambda) \bar{x}(\lambda) d\lambda \\ &= C(\omega_i, \omega_r) \int_\lambda \left\langle \left| P_{dtft} \left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda} \right) \right|^2 \right\rangle S(\lambda) \bar{x}(\lambda) d\lambda \\ &= C(\omega_i, \omega_r) \int_\lambda \left\langle \left| P_{dtft} \left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda} \right) \right|^2 \right\rangle S_x(\lambda) d\lambda \end{aligned} \quad (3.15)$$

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

3.1.3 Taylor approximation for BRDF

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

$$e^t = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \dots = \sum_{n=0}^{\infty} \frac{t^n}{n!} \quad (3.16)$$

Let us define

$$t = t(x, y) = ikwh(x, y) \quad (3.17)$$

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

$$\begin{aligned} e^t &= 1 + (ikwh) + \frac{1}{2!}(ikwh)^2 + \frac{1}{3!}(ikwh)^3 + \dots \\ &= \sum_{n=0}^{\infty} \frac{(ikwh)^n}{n!}. \end{aligned} \quad (3.18)$$

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:

$$\begin{aligned} \mathcal{F}\{p\}(u, v) &= \mathcal{F}\left\{\sum_{n=0}^{\infty} \frac{(ikwh)^n}{n!}\right\}(u, v) \\ &= \mathcal{F} \text{ lin Operator } \sum_{n=0}^{\infty} \mathcal{F}\left\{\frac{(ikwh)^n}{n!}\right\}(u, v) \\ &= \sum_{n=0}^{\infty} \frac{(ikw)^n}{n!} \mathcal{F}\{h^n\}(u, v) \end{aligned} \quad (3.19)$$

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

$$\sum_{n=0}^N \frac{(ikwh)^n}{n!} \mathcal{F}\{h^n\}(\alpha, \beta) \approx P(\alpha, \beta) \quad (3.20)$$

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

$$\forall y : \limsup_{k \rightarrow \infty} \left| \frac{a_{k+1}}{a_k} \right| = \limsup_{k \rightarrow \infty} \frac{y}{k+1} = 0 \quad (3.21)$$

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,

$$P_n(x) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x - a)^k \quad (3.22)$$

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:

$$|E_n(x)| \leq \frac{M}{(n+1)!} |x - a|^{n+1} \quad (3.23)$$

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

$$x_{max} = \|i\| k_{max} w_{max} h_{max} \quad (3.24)$$

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:

$$\begin{aligned}
x_{max} &= \|i\| * k_{max} * w_{max} * h_{max} \\
&= k_{max} * w_{max} * h_{max} \\
&= 2 * \left(\frac{2\pi}{4 * 10^{-7} m}\right) * 1.5 * 10^{-7} \\
&= 1.5\pi
\end{aligned} \tag{3.25}$$

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

$$\begin{aligned}
|E_n(x_{max})| &\leq \frac{M}{(n+1)!} |x_{max} - a|^{n+1} \\
&= \frac{\exp(1.5\pi) * (1.5\pi)^{n+1}}{(n+1)!}
\end{aligned} \tag{3.26}$$

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:

$$\begin{aligned}
X &= C(w_i, w_r) \int_{\lambda} \left\langle \left| P_{dtft}\left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda}\right) \right|^2 \right\rangle S_x(\lambda) d\lambda \\
&= C(w_i, w_r) \int_{\lambda} \left\langle \sum_{n=0}^N \frac{(wk)^n}{n!} \mathcal{F}^{-1}\{i^n h^n\}\left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda}\right) \right\rangle^2 S_x(\lambda) d\lambda
\end{aligned} \tag{3.27}$$

3.1.4 Sampling: Gaussian Window

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 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 $\mathbf{t}(\mathbf{x}, \mathbf{y}) = i^n h(x, y)^n$

$$\mathcal{F}_{dtft}^{-1}\{\mathbf{t}\}(u, v) = \mathcal{F}_{fft}^{-1}\{\mathbf{t}\}(u, v) window_g(\sigma_f) \quad (3.28)$$

Therefore we can deduce the following expression from this:

$$\begin{aligned} \mathcal{F}_{dtft}^{-1}\{\mathbf{t}\}(u, v) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_{fft}^{-1}\{\mathbf{t}\}(w_u, w_v) \phi(u - w_u, v - w_v) dw_u dw_v \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_i \sum_j F_{fft}^{-1}\{\mathbf{t}\}(w_u, w_v) \\ &\quad \delta(w_u - w_i, w_v - w_j) \phi(u - w_u, v - w_v) dw_u dw_v \\ &= \sum_i \sum_j \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_{fft}^{-1}\{\mathbf{t}\}(w_u, w_v) \\ &\quad \delta(w_u - w_i, w_v - w_j) \phi(u - w_u, v - w_v) dw_u dw_v \\ &= \sum_i \sum_j F_{fft}^{-1}\{\mathbf{t}\}(w_u, w_v) \phi(u - w_u, v - w_v) \end{aligned} \quad (3.29)$$

where

$$\phi(x, y) = \pi e^{-\frac{x^2+y^2}{2\sigma_f^2}} \quad (3.30)$$

3.1.5 Final Expression

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 2.7 the resulting color caused by the effect of diffraction can be computed like:
Let

$$P_\lambda(u, v) = F_{fft}^{-1}\{i^n h^n\}\left(\frac{2\pi u}{\lambda}, \frac{2\pi v}{\lambda}\right) \quad (3.31)$$

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

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = C(\omega_i, \omega_r) \int_{\lambda} \sum_{n=0}^N \frac{(wk)^n}{n!} \sum_{(r,s) \in \mathcal{N}_1(u,v)} |P_{\lambda}(u - w_r, v - w_s)|^2 \phi(u - w_r, v - w_s) \begin{pmatrix} S_x(\lambda) \\ S_y(\lambda) \\ S_z(\lambda) \end{pmatrix} d\lambda \quad (3.32)$$

where $\phi(x, y) = \pi e^{-\frac{x^2+y^2}{2\sigma_f^2}}$ is the Gaussian window 3.1.4.

3.2 Alternative Approach

3.2.1 PQ factors

In this section we are presenting an alternative approach to the previous Gaussian window approach 3.1.4 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:

$$S(x) = \sum_{n=0}^N f(x + nT) \quad (3.33)$$

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 3.27:

$$\begin{aligned} \mathcal{F}^{-1}\{S\}(w) &= \int f(x) e^{iwx} dx \\ &= \int_{-\infty}^{\infty} \sum_{n=0}^N f(x + nT) e^{iwx} dx \\ &= \sum_{n=0}^N \int_{-\infty}^{\infty} f(x + nT) e^{iwx} dx \end{aligned} \quad (3.34)$$

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

$$\begin{aligned} x &= y - nT \\ dx &= dy \end{aligned} \quad (3.35)$$

Plugging this substitution back into equation 3.34 we will get:

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

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:

$$\begin{aligned}
\sum_{n=0}^N e^{-uwnT} &= \sum_{n=0}^N (e^{-uwT})^n \\
&= \frac{1 - e^{iwT(N+1)}}{1 - e^{-iwT}}
\end{aligned} \tag{3.37}$$

We recognize the geometric series identity for the left-hand-side of equation 3.37. Since our series is bounded, we can simplify the right-hand-side of equation 3.37.

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

$$e^{-ix} = \cos(x) + i\sin(x) \tag{3.38}$$

Using the following trigonometric identities

$$\begin{aligned}
\cos(-x) &= \cos(x) \\
\sin(-x) &= -\sin(x)
\end{aligned} \tag{3.39}$$

combined with 3.38 we can simplify the series 3.37 even further to:

$$\frac{1 - e^{iwT(N+1)}}{1 - e^{-iwT}} = \frac{1 - \cos(wT(N+1)) + i\sin(wT(N+1))}{1 - \cos(wT) + i\sin(wT)} \tag{3.40}$$

Equation 3.40 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:

$$\begin{aligned} a &:= 1 - \cos(wT(N+1)) & b &= \sin(wT(N+1)) \\ c &= 1 - \cos(wT) & d &= \sin(wT) \end{aligned} \quad (3.41)$$

Hence, using 3.41, it follows

$$\frac{1 - e^{iwT(N+1)}}{1 - e^{-iwT}} = \frac{(a + ib)}{(c + id)} \quad (3.42)$$

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:

$$\begin{aligned} (cp - dq) &= a \\ (dp + cq) &= b \end{aligned} \quad (3.43)$$

After multiplying the first equation of 3.43 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 :

$$\begin{aligned} p &= \frac{(ac + bd)}{c^2 + d^2} \\ q &= \frac{(bc + ad)}{c^2 + d^2} \end{aligned} \quad (3.44)$$

Using some trigonometric identities and putting our substitution from 3.41 for a, b, c, d back into the current representation 3.44 of p and q we will get:

$$\begin{aligned} p &= \frac{1}{2} + \frac{1}{2} \left(\frac{\cos(wTN) - \cos(wT(N+1))}{1 - \cos(wT)} \right) \\ q &= \frac{\sin(wT(N+1)) - \sin(wTN) - \sin(wT)}{2(1 - \cos(wT))} \end{aligned} \quad (3.45)$$

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:

$$\begin{aligned} \mathcal{F}^{-1}\{S\}(w) &= \mathcal{F}^{-1}\{f\}(w) \sum_{n=0}^N e^{-iwnT} \\ &= (p + iq)\mathcal{F}^{-1}\{f\}(w) \end{aligned} \quad (3.46)$$

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 3.46.

$$\begin{aligned}
\mathcal{F}^{-1}\{S\}(w_1, w_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{n_2=0}^{N_1} \sum_{n_2=0}^{N_2} h(x_1 + n_1 T_1, x_2 + n_2 T_2) e^{iw(x_1+x_2)} dx_1 dx_2 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sum_{n_2=0}^{N_1} \sum_{n_2=0}^{N_2} h(y_1, y_2) e^{iw((y_1-n_1 T_1)+(y_2+n_2 T_2))} dy_1 dy_2 \\
&= \sum_{n_2=0}^{N_1} \sum_{n_2=0}^{N_2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(y_1, y_2) e^{iw(y_1+y_2)} e^{-iw(n_1 T_1+n_2 T_2)} dy_1 dy_2 \\
&= \sum_{n_2=0}^{N_1} \sum_{n_2=0}^{N_2} e^{-iw(n_1 T_1+n_2 T_2)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{Box}(y_1, y_2) e^{iw(y_1+y_2)} dy_1 dy_2 \\
&= \left(\sum_{n_2=0}^{N_1} \sum_{n_2=0}^{N_2} e^{-iw(n_1 T_1+n_2 T_2)} \right) \mathcal{F}^{-1}\{h\}(w_1, w_2) \\
&= \left(\sum_{n_2=0}^{N_1} e^{-iwn_1 T_1} \right) \left(\sum_{n_2=0}^{N_2} e^{-iwn_2 T_2} \right) \mathcal{F}^{-1}\{h\}(w_1, w_2) \\
&= (p_1 + iq_1)(p_2 + iq_2) \mathcal{F}^{-1}\{h\}(w_1, w_2) \\
&= ((p_1 p_2 - q_1 q_2) + i(p_1 p_2 + q_1 q_2)) \mathcal{F}^{-1}\{h\}(w_1, w_2) \\
&= (p + iq) \mathcal{F}_{DTFT}^{-1}\{h\}(w_1, w_2)
\end{aligned} \tag{3.47}$$

Where we have defined

$$\begin{aligned}
p &:= (p_1 p_2 - q_1 q_2) \\
q &:= (p_1 p_2 + q_1 q_2)
\end{aligned} \tag{3.48}$$

For the identity of equation 3.47 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 3.35, but this time twice, once for x_1 and once for x_2 separately. The last step in equation 3.47, substituting with p and q in equation 3.48 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 $(p^2 + q^2)^{\frac{1}{2}} |\mathcal{F}^{-1}\{h\}(w_1, w_2)|$

3.2.2 Interpolation

In 3.2.1 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 3.47 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:

$$x(t) = \sum_{n=-\infty}^{\infty} x[n] \text{sinc}\left(\frac{t-nT}{T}\right) \quad (3.49)$$

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.

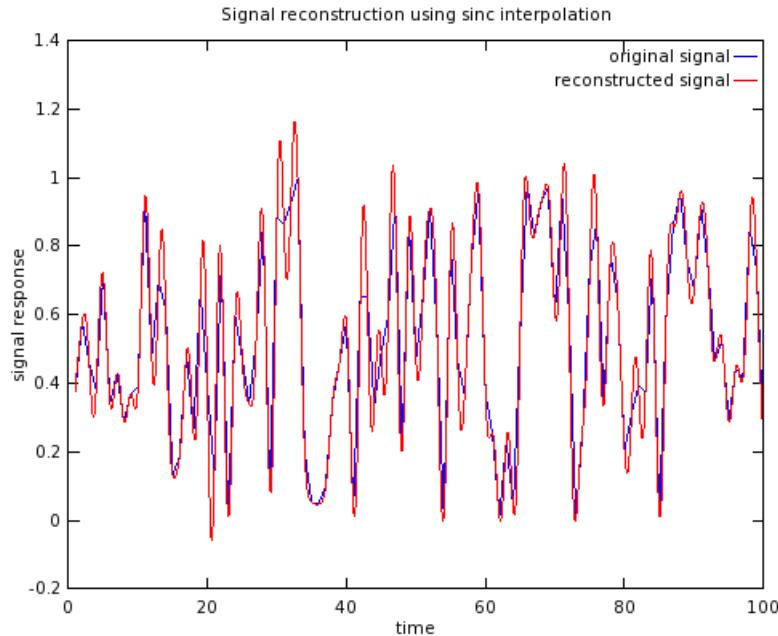


Abbildung 3.1: Comparission between a given random one dimensional input signal $s(t)$ and its sinc interpolation $\hat{s}(t)$. Notice that for the interpolation there were $N = 100$ samples from the original signal provided.

Anhang A

Appendix

A.1 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.1.1 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:

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

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:

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

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 $f(t)$.

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

where

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

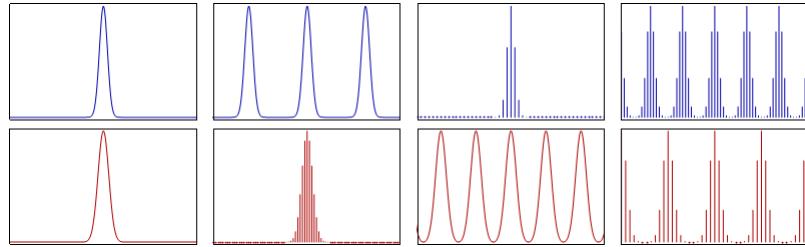


Abbildung A.1: Relationship¹ between the continuous Fourier transform and the discrete Fourier transform: Left column: A continuous function (top) and its Fourier transform A.1 (bottom). Center-left column: Periodic summation of the original function (top). Fourier transform (bottom) is zero except at 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.

Spacial 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.1.2 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.1.3 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.

¹image of illustration has been taken from wikipedia

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:

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

Anhang B

Appendix

B.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:

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

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.

B.2 Spherical Coordinates

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

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

B.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 B.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.

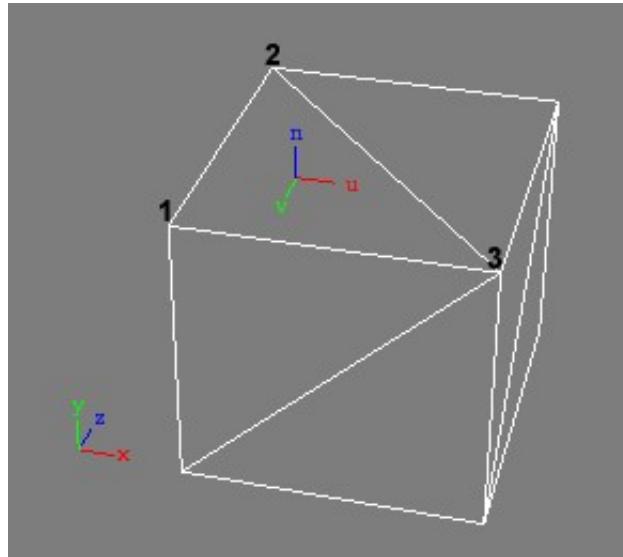


Abbildung B.1: Cube in world space (x, y, z) showing the tangen space (u, v, n) of its face (2, 1, 3)

Tabellenverzeichnis

A.1 Fourier operator to apply for a given spatial input signal and the properties of its resulting output signal in frequency space	30
---	----

Abbildungsverzeichnis

1.1	Effect of diffraction on snake sheds for different species	2
2.1	Irradiance is the summed up radiance over all directions	5
2.2	brdfillustration	7
2.3	visiblelightspectrum	8
2.4	humanayeschematic	8
2.5	Plots of our color matching functions we used for rendering	9
2.6	sinewave	11
2.7	ω_i points toward the light source, ω_r points toward the camera, n is the surface normal	13
3.1	Comparission between a given random one dimensional input si- gnal $s(t)$ and its sinc interpolation $\hat{s}(t)$. Notice that for the in- terpolation there were $N = 100$ samples from the original signal provided.	27
B.1	Cube in world space (x, y, z) showing the tangen space (u, v, n) of its face $(2, 1, 3)$	33

List of Algorithms

Literaturverzeichnis

- [Bar07] BARTSCH, Hans-Jochen: *Taschenbuch Mathematischer Formeln*. 21th edition. HASNER, 2007. – ISBN 978–3–8348–1232–2
- [DSD14] D. S. DHILLON, et a.: Interactive Diffraction from Biological Nano-structures. In: *XYZ* (2014), January
- [For11] FORSTER, Otto: *Analysis 3*. 6th edition. VIEWEG+TEUBNER, 2011. – ISBN 978–3–8348–1232–2
- [JG04] JUAN GUARDADO, NVIDIA: Simulating Diffraction. In: *GPU Gems* (2004). <https://developer.nvidia.com/content/gpu-gems-chapter-8-simulating-diffraction>
- [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
- [RW11] R. WRIGHT, et a.: *OpenGL SuperBible*. 5th edition. Addison-Wesley, 2011. – ISBN 978–0–32–171261–5
- [Sta99] STAM, Jos: Diffraction Shaders. In: *SIGGRPAH 99 Conference Proceedings* (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