ulm university universität
uulm

Simulation of Real Photographic Phenomena in Computer Graphics

- Revised version 1.1, October 2009 -

Diplomarbeit
Universität Ulm
Fakultät für Ingenieurwissenschaften und Informatik
Institut für Medieninformatik
Benjamin Steinert (mail@ben-steinert.net, 2009)

Erstgutachter: Prof. Dr. Hendrik Lensch, Universität Ulm
Zweitgutachter: Prof. Dr. Michael Weber, Universität Ulm

List of changes in version 1.1

- Added GPGPU aspects to future work.
- Convergence analysis removed.
- German *Eidesstattliche Erklärung* removed.
- Layout of Chapter 5 was revised.
- The following rendering results were replaced by improved versions:
  - Figure 1.1
  - Figure 6.10
  - Figure 5.5
  - Figure 5.18
  - Figure 6.4
  - Figure 6.5
  - Figure 6.3
  - Figure 6.7
- The following results were added afterwards.
  - Figure 5.9
  - Figure 6.11
- A few minor typos.

Abstract

The synthesis of photo-realistic images with today's techniques delivers high quality pictures with an outstanding degree of scene realism. However, an infinite field of focus, or a shot free of vignetting does not have much in common with a real photograph. In order to complete the chain of imperfect realism in computer graphics, light transport has to be simulated as if there was a real camera with a real lens mount and sensor. The need to merge real recorded image data and synthetically produced content without visible difference artifacts is one motivaton. Possible applications are found in movie industry, photography, virtual or augmented reality environments and digital image forensics.

The simulation environment is based on ray tracing methods considering the *Rendering Equation* in progressive stages by a *Monte Carlo Method* approach. A seperation of direct light from light sources provides an indedendent pass of rendering lens reflections. The presented method of sampling a pixel specific pupil for path generation at the sensor enables us to consider the correct effective aperture per pixel on the one hand, and minimizes the rate of blocked rays during passage on the other. A rudiment introduction of wave effects at the camera aperture by the *Geometrical Theory of Diffraction* and proposed sampling methods is given. We try to simulate diffraction effects in order to yield a basis for future experiments to receive results like the characteristic *Point Spread Function*. Implementation details propose concrete solutions of integration into an existing rendering system.

The thesis also provides interdisciplinary aspects of photography, introducing some practical elements for comparison and motivation. Results show that a simulation of all aberrations is absolutely possible.

Acknowledgments

First of all I would like to thank my mentors Prof. Hendrik Lensch, Prof. Michael Weber, Holger Dammertz and Johannes Hanika for the valuable support throughout writing this thesis. My additional correctors Katharina Schwarz, Sebastian Almer and Emanuel Simon had to revise many, really many mistakes, I'm glad to call you my friends. And I must not forget to thank my colleagues and friends in the student graphics lab for their valuable inspiration during writing this thesis and for all the other time doing weird off-topic stuff like photographing posters with black dots.

The scene models *Natural History Museum* and *Lighthouse* are freely availbale at `http://www.3drender.com`. The chess scene was published at `http://e2-productions.com`. The used meshes of the Stanford Dragon and Bunny, included in the CHROMA test scene, are provided by the Stanford Graphics Lab at `http://www-graphics.stanford.edu`. Thanks to the creators for their creativity and will to share their abilities with aristically rather untalented people like me.

Contents

List of Figures

List of Tables

List of Algorithms

1 Introduction

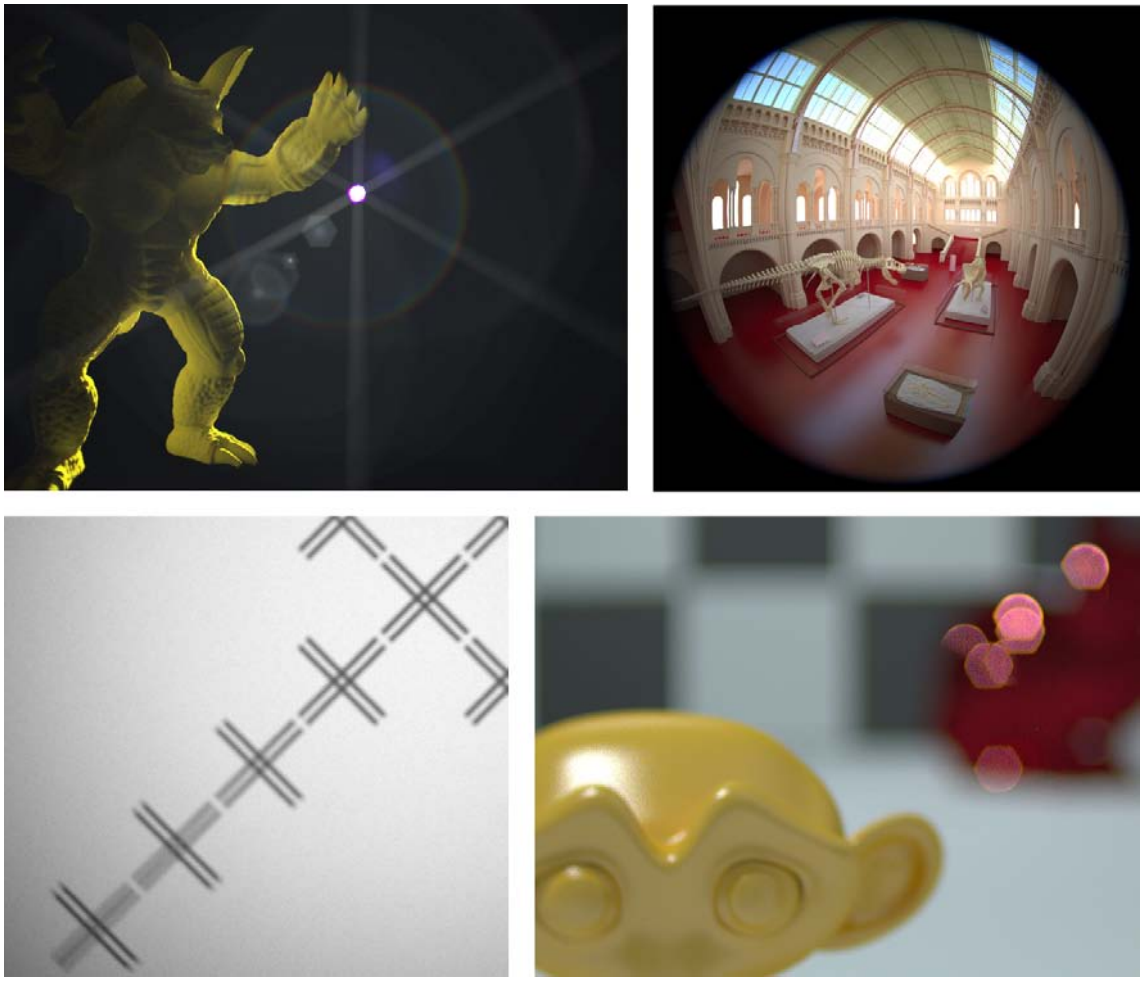


Figure 1.1: Example renderings, made with CHROMA, the prototype Monte Carlo path tracer developed during this thesis.

1.1 Motivation in Computer Graphics

The synthesis of photo-realistic images with today's techniques delivers high quality pictures with an outstanding degree of scene realism. However, an infinite field of focus, or a shot free of vignetting does not have much in common with a real photograph. In order to complete the chain of imperfect realism in computer graphics, light transport has to be simulated as if there was a real camera with a real lens mount and sensor. But such kind of model has often been discounted as far too expensive to simulate because approximations like the thin lens model, the pinhole camera or even post-processing techniques suffice the

needs of photographic effects in most cases. As soon as someone asks for effects like a correct lens flare or even chromatic aberration for example, the mentioned methods come to their limits.

Why should we introduce *errors* to a rendering technique, delivering better quality? The need to merge real recorded image data and synthetically produced content without visible difference artifacts requires to match both images including all imperfections of real cameras. Applications are found in film industry, photography, virtual or augmented reality environments and and digital image forensics. A recent example is Pixar's motion picture *WALL·E*. The german magazine *Digital Production* reported in an issue [Robo8] that Pixar redesigned the camera model of Renderman with a similar intent, in order to receive realistic imaging defects to make the resulting pictures more *dirty*. Such huge changes in a rendering pipeline, which demonstrated its ability in many projects before, underlines the important artistic component inherent to non-perfect images.

1.2 Aim of this Thesis

Towards receiving a physically based rendering system, able to accept a lens design together with appropriate parameters for glass materials, this work tries a balancing act between four different topics. The first considers optical physics, describing aspects like implied units, measures and the particular phenomena of light propagation. Further on the field of optical engineering gives an insight into how lens designs are described, which laws apply and what kinds of defects exist in an optical device. The third topic sets up the basis for the simulation environment and the tools that are available. Physically based image generation using ray tracing methods was chosen. The relatively intuitive models of light propagation apply best to that framework. Moreover, it gives the possibility to solve the *Rendering Equation* [Kaj86] by the *Monte Carlo Method*. Finally the fourth domain photography, itself already interdisciplinary, introduces some practical elements besides the technical aspects.

Inspiration for my thesis was the paper *A Realistic Camera Model for Computer Graphics* by Craig Kolb, Don Mitchell and Pat Hanrahan [KMH95]. The idea to use real lens designs in a distribution ray tracing setting evolved. I would like to extend their basic proposals, which mostly worked with approximate Gaussian methods originating from optical engineering context. Wavelength dependent effects and monochromatic aberrations are covered, Fresnel reflections inside the lens which cause flare effects are added and context to connected practical experience is given in several places.

Due to the complex character of light and the vast field of different topics, that has to be accounted for, several restrictions were made to the later simulation context. Polarization and phase, two properties of light, observed as wave, are neglected. However diffraction rudiments are added by sampling methods following the *Geometrical Theory of Diffraction* [Kel62]. Furthermore, time is not considered as influencing factor. Incidents at the light sensitive layer during exposure are not part of this thesis. Actually the chemical aspects of analog film or the behaviour of digital imaging arrays would fill another thesis.

2 Characteristics of Light

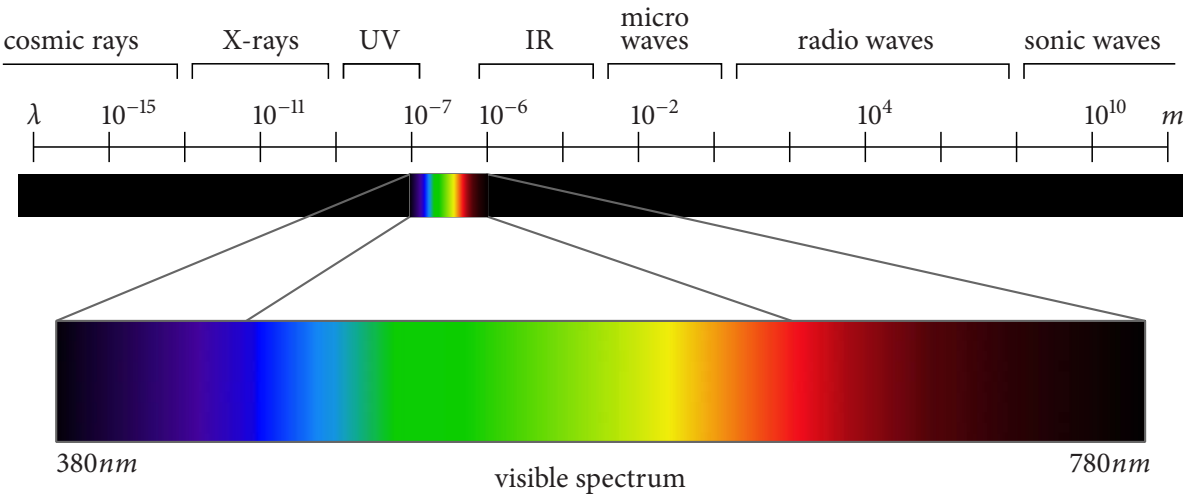


Figure 2.1: The electromagnetic spectrum with an approximate classification of notable regions.

This chapter gives an introduction on basic topics about light as part of the electromagnetic band. Propagation and surface interaction considerations are given, to be able to introduce these theories in later implementations. Not knowing, which specific units lie behind light in general, we start with the section on light duality and following properties.

Although visible light stays in the focus throughout the chapter, all rules hold for arbitrary wavelengths. The properties polarization and phase are mentioned in some place only to have an entry point for further reading or for the sake of completeness.

2.1 Geometric Optics

Considering light phenomena in general, the scientific field of optics is based on two different points of view. A first look at the electromagnetic spectrum in Figure 2.1 shows that this is indeed necessary. Cosmic and gamma radiation can be explained more easily by a particle approach, while radio or even acoustic spectra are much better understood in the wave-like way. Now the visible spectrum stands in the middle of these two points of view, which also means that both are applicable, but, from topic to topic, one is preferred [Smi07].

2.1.1 Light Propagation

The light propagation theorem by Fermat gives a fundamental law of how light travels in space:

The path a particle of light from point A to point B takes, has always the minimum optical path length. [PPBS05]

Consequently light travels along straight lines in a homogeneous medium, which allows us to choose rays as form of representation in this context.

2.1.2 Law of Reflection

An ideal reflection on a mirror surface If a light ray hits a mirror surface, we speak about ideal reflection. The incident direction and the surface normal form a plane in which also the reflection direction lies. The law of Reflection states:

$$\theta_i = \theta_r \tag{2.1}$$

as shown in Figure 2.2.

2.1.3 Law of Refraction

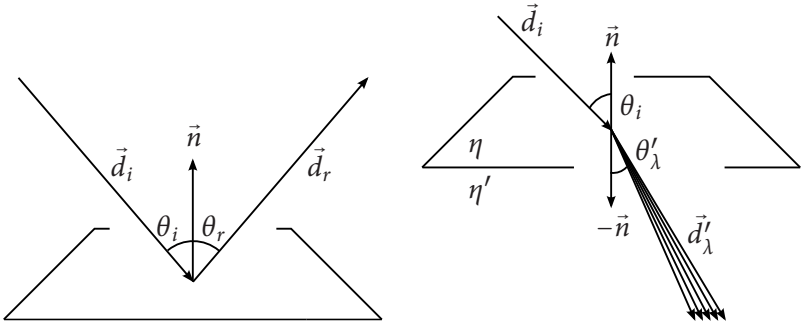
Light transmissive surfaces with a change in index of refraction from η to η' alter the travelling direction according to Snell's law (Figure 2.2):

$$\eta \sin\theta = \eta' \sin\theta' \tag{2.2}$$

A simple rules can be derived from that formula: During the transmission from the optically dense to a less dens medium, the light will be refracted away from the normal and vice versa. Note that refraction does not only result in transmission into the other medium, but also in reflection at the surface. Above the critical incident angle θ_c resulting in $\theta' = 90°$, reflection events according to (2.1) are received. From (2.2) and $\eta > \eta'$ we get

$$\theta_c = \arcsin\frac{\eta'}{\eta} \tag{2.3}$$

In practice, the index of refraction is not constant with the change of wavelength, which leads to the optical effect of dispersion, known from glass prisms. A ray of white light, consisting of different wavelengths, gets refracted in several rays different direction. In general, we can note that the lower the wavelength, the higher is the index of refraction [Smi07].



(a) Reflection at surface (b) Dispersive refraction at surface

Figure 2.2: Light interaction according to Fermat's Principle.

The amount of light travelling in the new direction through a refractive material is determined by the Fresnel equations [PPBS05]. For the special case of dielectrics, the generally complex valued index of refraction η stays real, which simplifies the equations and evaluation considerably. That is also the reason, why no absorption takes place, which is associated with the complex part of η. Consequently the amount of transmission T and reflection R sums up to one. The following equations only describe the transmission,

on the one hand perpendicularly polarized with $tr_\perp$ and on the other hand parallel polarized transmission $tr_\parallel$. Additional descriptions for reflection can be found in [BW75, PPBS05].

$$tr_\perp = \frac{2\eta\cos\theta}{\eta\cos\theta + \eta'\cos\theta'} tr_\parallel = \frac{2\eta\cos\theta}{\eta'\cos\theta + \eta\cos\theta'} \tag{2.4}$$

And finally

$$T_{\perp\vee\parallel} = \frac{\eta'\cos\theta'}{\eta\cos\theta} tr^2_{\perp\vee\parallel} \tag{2.5}$$

Averaging both values can yield a solution without consideration of polarization.

2.1.4 Absorption, Transmission and Scattering

Absorption is explained by surfaces that are not able to re-emit incident light of certain wavelengths. For example a white piece of paper reflects all wavelengths, but a red piece only the ones with $\lambda > 600nm$. The missing part is conducted in form of heat.

During the transmission through glass media, light intensity gets reduced by certain impurities and so-called *body colour* [Ray02], which gives glass a certain colour. These properties define an absorption coefficient α which relates incoming and dielectric travelling light for a transmission distance t:

$$\begin{aligned} L_t &= L_i\sigma_b \\ \sigma_b &= \exp(-\alpha t) \end{aligned} \tag{2.6}$$

This law is called Beers Law or Bougher Lambert Absorption [Ray02].

Scattering is the effect when light particles collide with other particles, like in dust clouds. Due to the fact that scattering events have almost no impact in clean air, it can be neglected so scattering is only mentioned for the purpose of completeness and will not be further needed here.

2.2 Diffraction

Diffraction phenomena are widely described as a wave effect in optics, because they are not explainable with the classic principle by Fermat [PPBS05]. To become familiar with the incidence of diffraction and the thought behind wave optics, I will explain both briefly before I will return to a ray approach based on Geometrical Theory of Diffraction.

2.2.1 Huygens' Principle of Light Propagation

Diffraction is the optical phenomenon that light can travel around edges of obstacles. That can be easily shown by a typical photographic setting like in Figure 2.3, where edges cause very high contrast between enlightened and dark regions.

The classic experiment of proof in the lab would be done with a circular aperture or a slit in front of a screen. Light which passes obstacles, that create shadow boundaries, does not form expected hard borders between shadow and lighted region. That incident contradicts Fermat, because light seems to be bent around the edge. Results that reveal interference can be gained with similar experiments, but are left out in this place.

Figure 2.3: Photographs directly into the sun are feared due to the resulting, most times unwanted, optical effects. But they can also be beautiful. The streaks along the twigs which form the obstacle, consist of diffracted light, scattered towards the viewing direction.

The dutch physicist Christiaan Huygens proposed the wave principle of light propagation, which was able to explain all laws resulting from Fermat's' principle, and, in addition, the diffraction effect at obstacles [PPBS05, BW75]. Light is travelling in the form of wave fronts with a certain wavelength and amplitude. Assuming an isotropic medium, the wave fronts emanate spherically from a point light source. That alone does not yet result in diffraction, it is just another point of view, if we think of waves as perpendicular lines to our rays from the Fermat point of view. The following sentence describes the fundamental property of light waves:

Every point on an emanating light wave is origin for a further wavelet of same frequency, wavelength and polarisation, whose envelope again implies a new wave front. [BW75, PPBS05]

This law also defines the wavelets at edges of the aperture and which the light has to pass. Because of the spherical propagation, there exist wave fronts behind obstacles that normally form the shadow boundary. Figure 2.4 visualises the consequence in detail.

2.2.2 Geometrical Theory of Diffraction

For later purpose, the wave model of light is not suitable. Joseph B. Keller published a new approach to diffraction in the Fermat way of view [Kel62]. For that he proposed several modified theorems, which account for specific geometric settings. The most important one concerns diffraction on a straight edge:

An edge-diffracted ray from a point P to a point R is a curve which has stationary optical length among all curves from P to R with one point Q on the edge.

The edge itself is part of a wedge with opening angle α where

$$\alpha = \pi(2 - n). \tag{2.7}$$

But as we already know, diffraction does not allow for sending the ray further in only one deterministic direction. All valid diffraction directions are rays which lie perpendicular to a chosen wave emanating from the point of diffraction. Keller's results concentrate on the introduction of the so-called diffraction cone shown in Figure 2.5. Its surface describes all valid directions for diffracted rays in the diffraction point Q. The angle of incidence β to the edge vector equals the half opening angle of the cone. A ray with normal

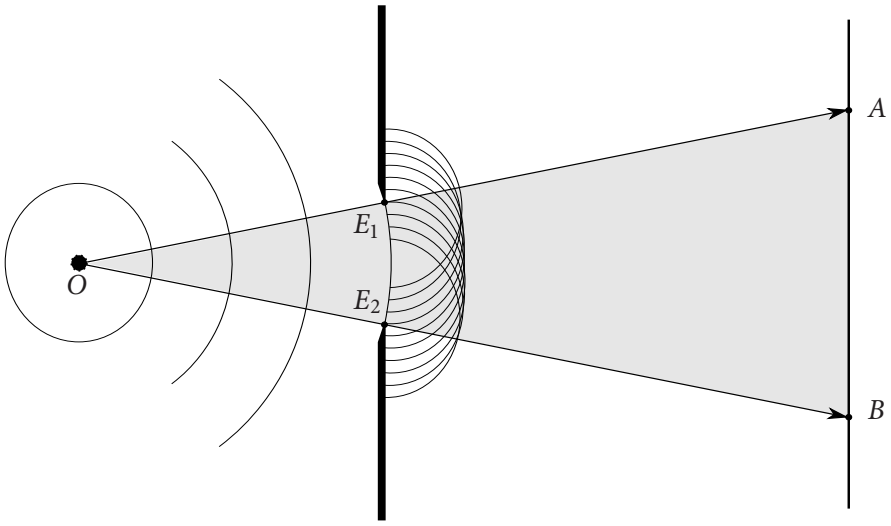


Figure 2.4: Light originating from point O through the slit with edges in E_1 and E_2. With Fermat's' principle the lit region lies between A and B. In fact the light reaches areas far beyond those points, it gets even reflected back at the point of diffraction, not drafted here for the sake of clear arrangement.

incidence stills holds for that rule. Result is just a cone with length zero, or simply a perpendicular disc around the diffraction point Q.

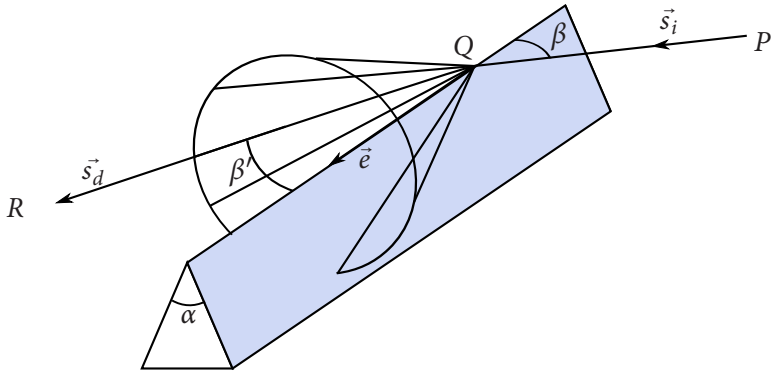


Figure 2.5: The cone of diffracted rays with diffraction edge as axis and opening angle $\beta = \beta'$ stated by Keller. The edge is settled on a wedge with inner angle α.

When studying the two dimensional example in Figure 2.6 we rediscover one already known incident, the law of reflection. The outline decomposes the electromagnetic field at a point R in its contribution fields given by the setting. E_i can be seen as differential tubes of the incident wave front, formed by rays travelling through the region A and B directly. Additionally region A gets indirectly illuminated by reflection at the geometry. Region C would not be illuminated under standard geometric optics considerations. By diffracted field rays E_d, all locations in all regions receive contribution in the two dimensional case. The incident angle β is always 90° here, which leads us to the special case of having the valid diffraction directions on the just mentioned perpendicular disc.

The the diffracted field at R can be obtained by

$$E_d(R) = E_i(Q) \cdot D \cdot A(s) e^{-jks} \tag{2.8}$$

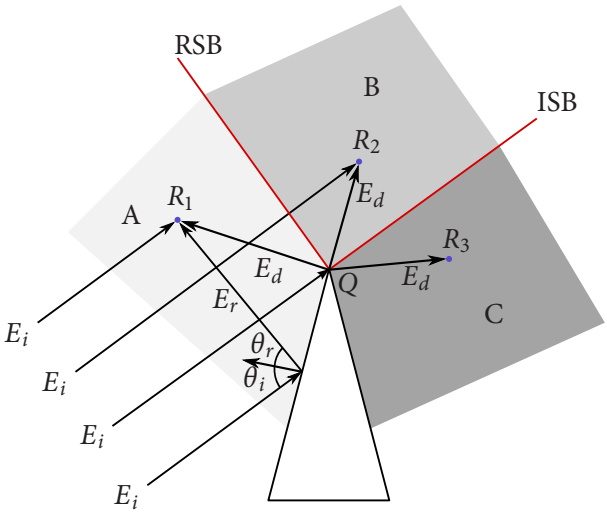


Figure 2.6: Definition of areas with different electromagnetic field contributions E_*. Region A is illuminated directly, by reflection and by diffraction. Area B is outside the reflection zone, and part C is only illuminated by diffracted light. RSB stands for reflection shadow boundary, and ISB for illumination shadow boundary respectively.

where s is the distance between Q and R, D is the diffraction coefficient, $k = \frac{2\pi}{\lambda}$ and $A(s)$ introduces the spreading factor given by the wave front form and edge shape. For spherical waves hitting a straight edge

$$A(s) = \frac{1}{\sqrt{s}}. \tag{2.9}$$

To state the amplitude and phase of diffracted light, there exist different diffraction coefficients derived by Keller and generalized by Kouyoumijan and Pathak [MPM90]. But before we get stuck too deep in the rather unaccustomed theory, I leave them out at that point of explanation, because the general formulae contain consideration of phase and polarization. For the original proposals of Keller please consult [Kel62].

At this point I would like to leave it at that, because further consideration of theory does not add more knowledge for the later introduced adaption to computer graphics. What should be kept in mind from this section, is the general point of view on waves and the thoughts on diffracted ray propagation based on the Keller cone. For deeper explanations of the models and theories behind it, I strongly recommend to go through [Kel62] for the first idea, which can be broadened by consulting [MPM90].

2.3 Radiometry

The science of Radiometry applies to the whole spectrum, from gamma radiation to acoustic waves, and only handles physically measurable units. The more psychological science of Photometry only considers the for humans visible part of the spectrum from approximately 380nm to 770nm [PPBS05]. Its base of existence is the fact that light with the same intensity and different wavelength is perceived unequally *bright*.

Geometric optics in general assumes light to consist of particles called photons, with frequency f and varying wavelength λ. Planck stated that one such photon carries the specific energy

$$q = \frac{hc}{\lambda}\,[J] \tag{2.10}$$

with Plank's Constant $h = 6.63 \cdot 10^{-34} Js^{-1}$ and light velocity constant $c = 299792458 ms^{-1}$.

But in general, light is not regarded as a count of photons but as a certain continuous field of different wavelengths. The energy Q of a specific spectrum from $\lambda_1$nm to $\lambda_2$nm is defined by:

$$Q = \int_{\lambda_1}^{\lambda_2} Q_\lambda \, d\lambda. \tag{2.11}$$

Radiant flux or power Φ introduces time t

$$\Phi = \frac{dQ}{dt} \left[\frac{J}{s}\right] = [W]. \tag{2.12}$$

The flux traveling in a certain direction in steradians[1] is named radiant intensity with

$$I = \frac{d\Phi}{d\omega} \left[\frac{W}{sr}\right]. \tag{2.13}$$

For example stars in astronomy can be approximated by thinking of a point radiating Φ watts equally in all directions, which means that the radiant intensity is $\frac{\Phi}{4\pi} \left[\frac{W}{sr}\right]$ That is the classical idea of a hypothetical point light source.

The power reaching a surface patch is called irradiance

$$H = \frac{d\Phi}{dA} \left[\frac{W}{m^2}\right]. \tag{2.14}$$

To continue our example, imagine a photovoltaic cell on a satellite. The irradiance onto the surface with its normal intersecting the center of the star is now given by the fraction of radiant intensity of the star and the solid angle subtended by the cell.

Here we can introduce the essential inverse square law [PPBS05], which describes the correlation between irradiance and distance. Now go one step further and think about a gigantic sphere[2] with radius r centered around our star. Hereby we subtend 4π sr seen from the star location. Now, by (2.13) and (2.14)

$$H = \frac{d\Phi}{dA} = \frac{4\pi I}{4\pi r^2} = \frac{I}{r^2} \left[\frac{W}{m^2}\right]. \tag{2.15}$$

resulting in a direct dependence of irradiance from distance to the source, which is obvious after studying figure Figure 2.7a

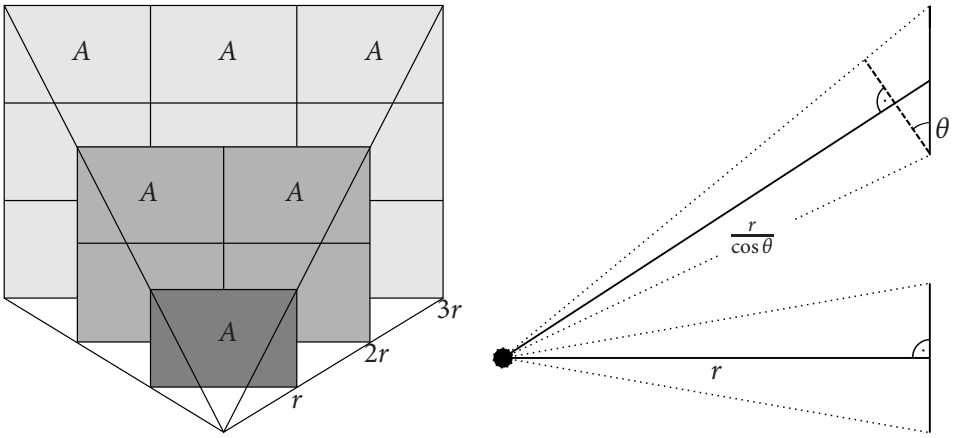
Let's come back to our cell which is exactly perpendicularly oriented to the normal through the star in only one point! With increasing distance to that point, the angle the angle of incidence varies. The projected area is reduced by the cosine of the angle differing from the normal direction and we can introduce a variation with the movement perpendicular to the normal. As seen in Figure 2.7b the distance differs with $\cos\theta$, resulting in a $\cos^2\theta$ due to (2.15) and $\cos\theta$ for resizing the projected area.

Now we leave our star and come to more practical light sources with an areal extension. Here we will speak about emitted radiance L concerning the energy leaving an area differential on the source in one direction.

$$L = \frac{dI}{dA\cos\theta} = \frac{d^2\Phi}{d\omega dA\cos\theta} \left[\frac{W}{m^2 sr}\right]. \tag{2.16}$$

[1] steradians, or sr, is the derived SI unit for solid angle. Its value describes the size of a piece of area projected onto a sphere around the point of interest. For example a sphere subtends 4π sr from its center

[2] the theoretical model indeed already exists and is called *Dyson Sphere*



(a) Irradiance reduces by a factor of four with doubled distance.

(b) Irradiance onto a surface with angle θ to the light direction which reduces projected area by $\cos\theta$. Shifting the position of the surface increases distance r by $\frac{1}{\theta}$. Overall result is an approximate $\cos^3$ variation of irradiance with θ for large distances.

Figure 2.7: Variation of irradiance with distance and angle.

The attenuation by $\cos\theta$ is needed because the intensity value is only valid for direction along the normal direction. With increasing angle, the projected area dA approaches zero. Irradiance can be reintroduced by integrating over the hemisphere domain, written S^2_-. In non-mathematical words, we *collect* the radiance over all directions.

$$H = \int_{S^2_-} L(\omega)\cos\theta \, d\omega \tag{2.17}$$

When we speak about light source emission on a surface in the following, emitted radiance L_e and radiant exitance E are the correct terms of use. On the other hand, incident radiance L and irradiance H from a light source onto an area are the terms for the contrary point of view.

2.3.1 Case study: The Lambert Emitter

A typical model of a light source is the Lambert emitter. Like the name says, the emission behaviour is lambertian or ideal diffuse, meaning that

$$I(\theta) = I_0 \cos\theta \tag{2.18}$$

where I_0 is the radiant intensity in normal direction. Now with equation (2.16), we can derive a fundamental consequence: The radiance of a lambertian light source is constant with respect to θ [Smi07].

$$L(\theta) = \frac{I(\theta)}{A\cos\theta} = \frac{I_0\cos\theta}{A\cos\theta} = \frac{I_0}{A} = const. \tag{2.19}$$

(2.17) can now be simplified to:

$$E = \int_\Omega L(\omega)\cos\theta \, d\omega = L\int_\Omega \cos\theta \, d\omega = L\int_0^{2\pi}\int_0^{\frac{\pi}{2}} \cos\theta\sin\theta \, d\theta d\phi = \pi L \tag{2.20}$$

Note that the result is not $2\pi L$ as one could expect at first sight when thinking of radiance emitted into a hemisphere with $2\pi sr$. One π gets lost due to the area fall-off with respect to the pole angle θ.

2.3.2 Surface Reflectance

The reflectance behaviour of a surface can be generally described by the bidirectional reflectance distribution function $\rho(\omega_i, x, \omega_o)$ short BRDF. It can either have an analytic description or is a discrete table as result of experiments with a gonioreflectometer[3]. The most important property is, that it observes the law of energy conservation. The easiest analytic form is the Lambertian BRDF, directly related to the Lambert emitter. The result is an equally reflecting surface over all directions ω_o in the normal-side hemisphere S^2_- of the received irradiance.

$$\rho(\omega_i, x, \omega_o) = \frac{L(\omega_o)}{H} \tag{2.21}$$

For a diffuse surface with reflectance R and constant BRDF ρ, reflected radiance is constant over the hemisphere (compare with Lambert emitter). We get

$$\begin{aligned} R &= \int_{S^2_-} \rho \cos\theta_i \, d\omega \\ &= \rho \int_{S^2_-} \cos\theta_i \, d\omega \\ &= \rho\pi \\ R &\leq 1 \rightarrow \rho \leq \frac{1}{\pi} \end{aligned}$$

Be aware, that reflectance, not the BRDF, must fulfill the law of energy conservation to be ≤ 1 because it is applied to the full hemisphere. Otherwise the surface would have a reflectance of π.

A second, at first sight simple BRDF seems to be the ideal mirror reflection. By usage of the Dirac pulse concept $\delta(0) = 1$ an analytic description can be formulated:

$$\begin{aligned} L(\omega_0) &= \int_{S2-} \rho(\omega_i, x, \omega_o) L(\omega_i) \cos\theta_i \, d\omega \\ &= \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \rho((\theta_i, \phi_i), x, (\theta_o, \phi_o)) L((\theta_i, \phi_i)) \cos\theta_i \sin\theta_i \, d\theta_i d\phi_i \\ &= \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \frac{\delta(\theta_i - \theta_o)}{\cos\theta_i} L((\theta_i, \phi_i)) \cos\theta_i \sin\theta_i \, d\theta_i d\phi_i \\ &= L(\omega_i) \end{aligned}$$

2.3.3 Wavelength Revisited

In the beginning of this section we started with the integration over a certain spectrum in equation (2.11). Once we want to work with a specific wavelength we can do so, by reintroducing wavelength dependency in our formulae. The radiometric signs become subscripted with a specific λ and their names get prefixed *spectral*. For example equation (2.16) in context of spectral dependence becomes:

$$L_\lambda = \frac{dI_\lambda}{dA\cos\theta} = \frac{d^2\Phi_\lambda}{d\omega dA\cos\theta} \left[\frac{\frac{W}{nm}}{m^2 sr} \right]. \tag{2.22}$$

[3] a device which measures reflected radiance for all directions according to incident radiance from all directions

To keep confusion minimal with the resulting m^3 in the denominator, maintain the fraction of $\frac{W}{nm}$, or use the alternative SI unit of Ångström $1Å = 10^{-10}m = 0.1nm$.

The definition of how much contribution a certain wavelength has for an emitter is described by its spectral power distribution, which gives relative or absolute values for a specific radiometric unit like radiant exitance for example.

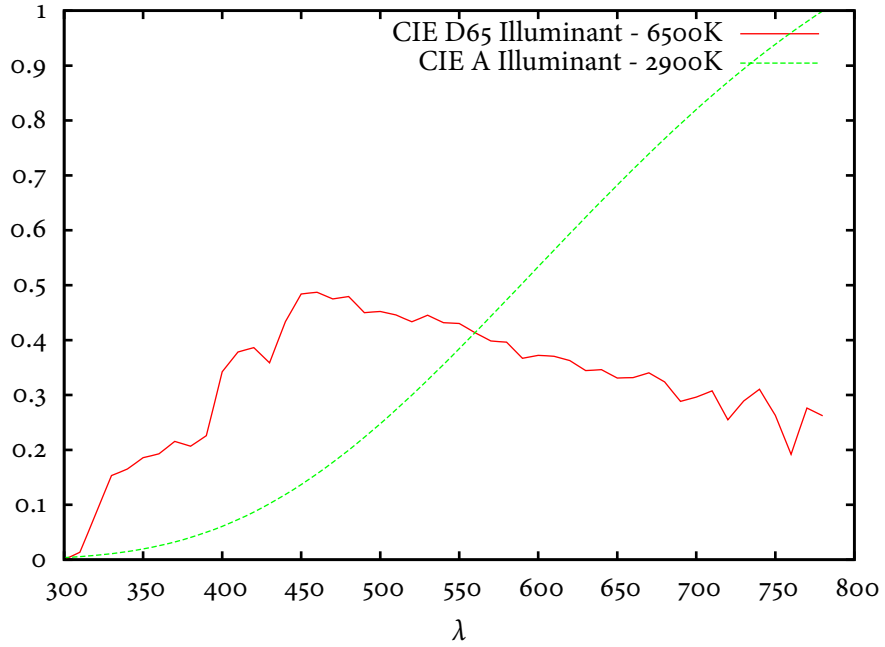


Figure 2.8: Plot of two spectral power distributions for CIE standard light sources. [CIE04]

This function is obviously also responsible for the color of the light source which is expressed in the temperature unit Kelvin (K). That probably unexpected description has its origin in the traditional description of the Planck black body radiator. This type of emitter has an analytic form of spectral power distribution, called Planck's law [PPBS05, Smi07]:

$$E_\lambda = \frac{2\pi hc^2}{\lambda^5\left(e^{(hc)/(\lambda Tk)} - 1\right)} \tag{2.23}$$

Inserting Planck's constant h, light velocity c and the Boltzmann constant k yields

$$E_\lambda = \frac{3.742 \cdot 10^8}{\lambda^5\left(e^{14388/(\lambda T)} - 1\right)} \quad \left[\frac{W}{m^2 \mu m}\right] \tag{2.24}$$

where T stands for a temperature in Kelvin, and λ has unit μm. This law is not applicable to normal light sources, easily recognizable by the tremendous temperatures, but the visual appeal is considered for comparison. When a light source is said to have 6500K, it is meant, that the light has the same perceived colour than a 6500K black body emitter. This very inaccurate definition has been established as colorimetric description, and is not capable of replacing a correct spectral power distribution.

2.4 Photometry

When we speak of photographic quantities, we have to adapt the radiometric values to the perception of the human eye. All radiometric units are renamed, and the symbols change again. In this thesis, they

become subscripted with v, to keep the mix-up somehow minimal. The cognition of colour of our eyes and

symbol	radiometric unit	photometric unit
Q	Joule [J]	Lumen seconds[lms]
Φ	Watt [W]	Lumen [lm]
I	$\frac{W}{sr}$	Candela [cd]=$\left[\frac{lm}{sr}\right]$
H,(E)	$\frac{W}{m^2}$	Lux [lx]=$\left[\frac{lm}{m^2}\right]$
L	$\frac{W}{m^2 sr}$	Candela per $m^2\left[\frac{cd}{m^2}\right]$

Table 2.1: Radiometric and photometric units.

brain is not linear with wavelength, which introduces the so called luminous efficiency functions $V(\lambda)$ and $V'(\lambda)$ [PPBS05]. They differ in the way, the human eye works. In detail, there exist two types of cells in the eye, one for day vision, the so-called cones, and the rods for night vision. Figure 2.9 illustrates the sensitivity depending on wavelength.

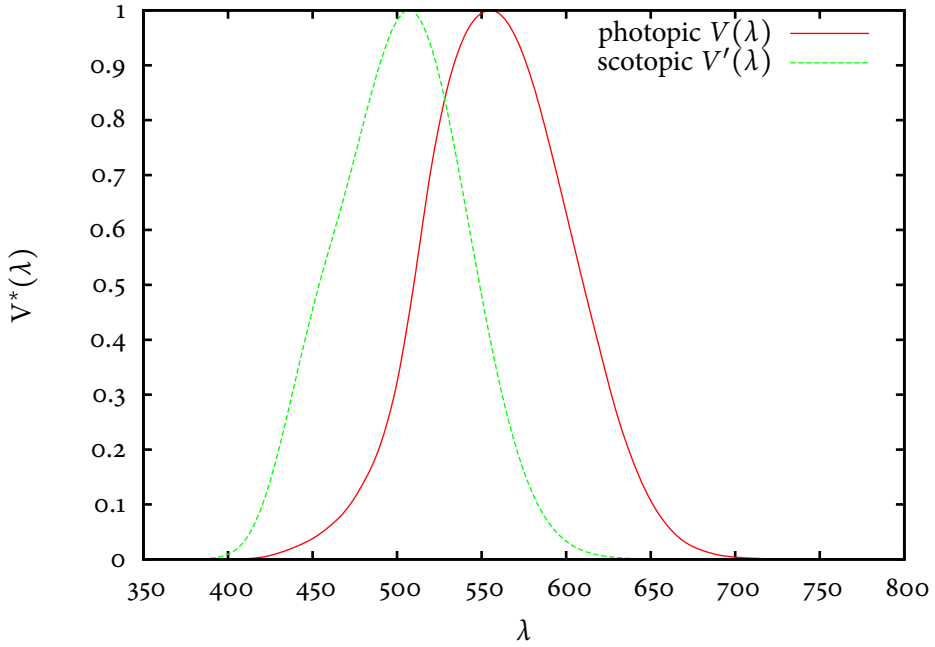


Figure 2.9: Luminous efficiency of the humans' eye receptors

To derive a photometric quantity P from its spectral radiometric equivalent R_λ we use:

$$P = C\frac{lm}{W}\int_{\lambda_1}^{\lambda_2} V(\lambda)\, R_\lambda(\lambda)\, d\lambda \tag{2.25}$$

with $C = 683$ for photopic day vision and $C = 1699$ with $V(\lambda)$ replaced by $V'(\lambda)$ for scotopic night vision.

This is one reason for the visual perception of different spectra to be the same color, which are then called metamers. A change in the distribution aside the important area around 555nm results in less influence to the appropriate eye receptors. That is also the reason why it is so significant to have a high value around 555nm to receive the maximum perceived brightness according to consumed power. Colorimetry suffers from metamerism in further situations. Every finite dimensional color definition has an infinite number of spectral representatives that express the same color.

3 Still Camera Technology

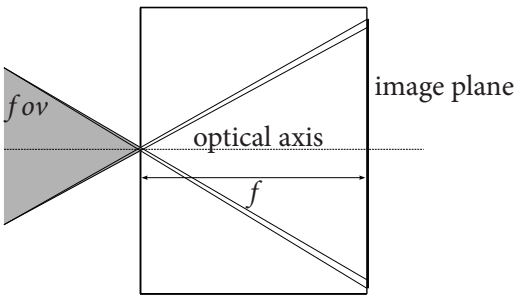
The well established science of optical engineering evolved a wealth of techniques in the past to form *light information* for arbitrary viewing conditions. Chemistry in the 19th set the the ball rolling for permanent records.

First of all I briefly introduce two similar types of constructions, with the one distinction of using lens elements for image formation. The introduction of the latter yields several new problems together with its possibilities, also explained in this chapter.

3.1 Camera Models

3.1.1 Camera Obscura

(a) Modern version with 0.22mm diameter hole in 0.02mm thick plate mounted on DSLR back

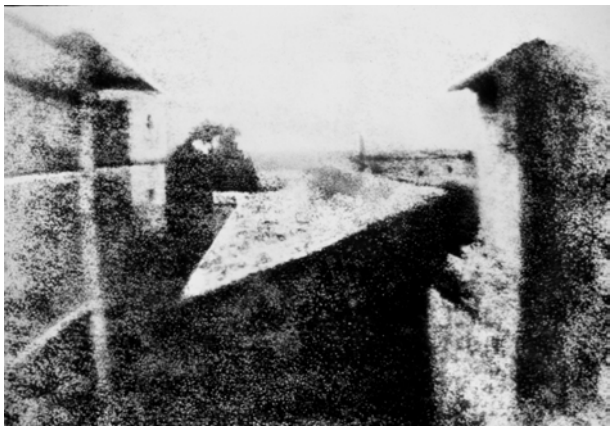


(b) Pinhole camera schematic

Figure 3.1: Modern versions of pinhole cameras are relatively easy to build. A body cap was used to insert a thin plate where the hole was eroded into.

A photographic camera is a device with the purpose of recording an image by a light sensitive layer. To control the angle and amount of light falling in, there exist a variety of extensions to that general apparatus. The term “photograph” originates from the Greek words “photos” and “graphos, meaning “light” and “drawing”.

The oldest form of a camera as an imaging system is the pinhole camera or *camera obscura* [Per07]. First written mentions are dated back to the 5th century, with the typical method of operation. Light falls through a very small hole into a dark box and projects an upside down image of the scene onto the image plane. First real photographic emulsions to record images on certain materials were introduced in the beginning of the 19th century, when advances in chemistry yielded substances with light-sensitive character, fast enough to record an image under useful settings. Before that, the image forming principle was used a long time for example by artists to project a motif onto a canvas as a sketching assist.



(a) (1826) First confirmed permanent camera obscura lithography by J. N. Niépce.

(b) (2009) Picture of the *Ulmer Muenster*: Left image taken with an equivalent pinhole camera mount shown in figure 3.1. The right one is taken with a professional Canon lens.

Figure 3.2: Pinhole photography examples

The french inventor Joseph Nicéphore Niépce succeeded around 1826 to create the first positive photograph with a camera obscura [Per07]. The exposure time is the controlling factor of picture brightness, under consideration of the pinhole diameter [PPBS05, Ray02]. The field of view is based on the distance between film and the so-called center of projection. It is also to mention, that a pinhole photograph has an infinite

depth of focus, while the image sharpness mainly depends on the pinhole diameter forming the circle of confusion. That is the image size of a point, like drafted in Figure 3.1b, by the diverging to lines from the same origin through the aperture. The circle diameter is approximately constant, for relatively large image distances, compared to the hole radius.

Today, pinhole photography is still an asked technique for photographic art, although the possibilities are limited due to the relatively long exposure times from seconds to hours.

3.1.2 Lens Camera

The completely independent science of optical instrument design is obviously much older, than photography as we know it. Telescopes and viewing glasses were already vastly known and used. To overcome the problem of long exposure times, one has to bring more light onto the sensitive layer, without the loss of image quality. Already in the 16th and 17th century, various improvements on the pinhole camera model were made, introducing a biconvex lens in front, a diaphragm to control the pinhole diameter and a 45 degree reflex mirror, to project the image onto an alternative plane [Per07]. After the death of Niépce in 1833 his inventions and ideas about light sensitive materials stood unpublished. His son Isadore continued the work together with Louis Daguerre with whom his father started to cooperate in 1829. One of Daguerre's supporters was Charles Chevalier a leading optician of that time. His optical knowledge and the approaches of Daguerre and Niépce resulted in the successful completion of the Daguerrotype which was presented to the public in 1839. Their invention is taken as basis for today's photographic cameras [Ray02] although most of the basic concepts were already established by other inventors far earlier. Their main contribution was the revolutionary short exposure time they needed for an image.

In general, by combination of optical engineering and photography, the amount of image forming factors is extended by a bunch of new complex possibilities and problems which are now explained in the following section.

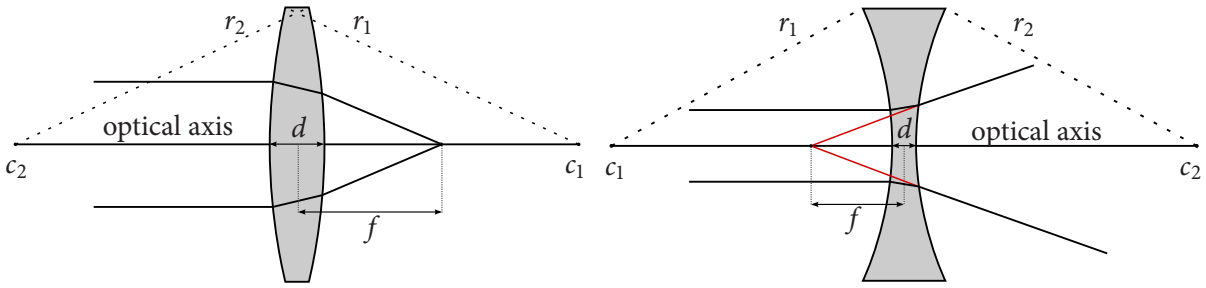
3.2 Photographic Optics

3.2.1 Introduction

Optical instruments built in the classic way, consist of one or more spherically curved, axial symmetric glass surfaces [Ray02] which refract light according to the laws mentioned in Section 2.1. The curvature of surface and the material itself define the optical behaviour. A convex lens is also called converging lens, because it concentrates or converges light to a focus point. A concave lens works the other way round, it diverges light, seen from a hypothetical focus point on the incident side. Combinations of convex and concave surfaces are called meniscus with the prefix positive or negative, depending on whether they are mainly converging or diverging. The optical power of a lens is given by the Lensmaker's equation

$$F = 1/f = \frac{\eta' - \eta}{\eta}\left(\frac{1}{r_1} - \frac{1}{r_2} + \frac{(\eta' - \eta)d}{\eta r_1 r_2}\right) \quad [diopters] = \left[\frac{1}{m}\right] \tag{3.1}$$

where r_1 is the radius of curvature of the surface nearer to the emanating point, r_2 the farther one. Variable d is the lens thickness along the optical axis, η' is the refraction index of the material, η the index for the surrounding substance. There is an important rule when considering the radii of curvature described in an optical engineering context:



(a) Converging lens (b) Diverging lens

Figure 3.3: Basic spherical lens forms

A positive radius stands for a surface with center of curvature on the right side or curved convex when viewed from the left, and negative sign means that the center is on the left side, respectively seen concave from the left [Smio7].

A usage of lenses in tandem, either cemented or separated, gives further possibilities, especially degrees of freedom for aberration correction.

3.2.2 Aberration Theory

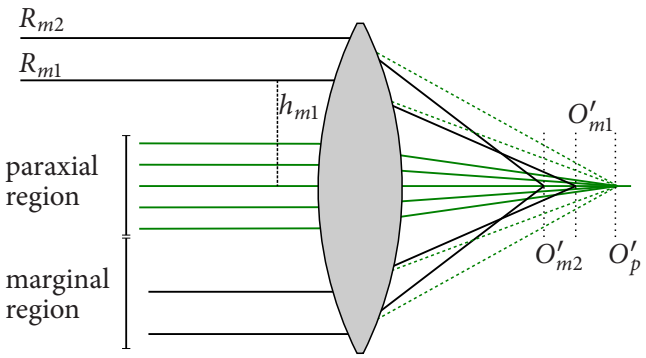


Figure 3.4: Exaggerated comparison of Gaussian refraction (dashed) and Snell's refraction (solid) of marginal region rays.

This section goes back to a more optics theory related topic. We have discussed lenses that focus incoming light to a point, but that is not true, obviously after studying figure 3.4. The varying thickness and incident angle due to the spherical surface cause a changing optical path length. Consequently the image suffers defects called Seidel aberrations, named after the mathematician Ludwig von Seidel who first classified the different types [Rayo2, Smio7].

Tracing rays through an optical system means applying Snell's law (2.2) repeatedly:

$$\eta \sin\theta = \eta' \sin\theta'$$

An expansion of $\sin\theta$ by the Taylor series reveals, that the incident angle comes up in different orders.

$$\sin\theta = \sum_{n=0}^{\infty} (-1)^n \frac{\theta^{2n+1}}{(2n+1)!} = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} \pm \cdots \tag{3.2}$$

When discarding all terms higher than order of one, we get

$$\sin\theta = \theta\,(+O(n^2)) \quad [rad] \tag{3.3}$$

and come to the paraxial law of refraction

$$\eta\theta = \eta'\theta'$$

which is mentioned for comparison with reality. It is the basis for the field of *Gaussian Optics* [PPBS05, Smi07], a very versatile analysis tool in geometric optics, but which has nothing to do with reality, as soon as the higher order terms have a certain impact [Smi07, Ray02]. The truncated higher order terms in equation 3.2 introduce aberrations of different order varying with the height of the paraxial ray[1]. Figure 3.4 underlines the difference between Gaussian optics and reality for marginal rays with large height h.

Seidel classified five different forms of aberration introduced with the 3rd order. They are are noted as monochromatic, meaning that they are independent of wavelength in opposition to chromatic aberrations, also explained later.

Spherical Aberration

Spherical Aberration is almost explained with Figure 3.4. But for the sake of completeness and detailed graph descriptions a modified version can be found in Figure 3.5

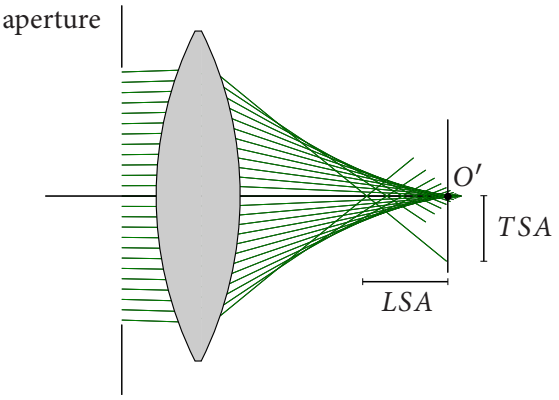


Figure 3.5: Spherical aberration: Can be described either longitudinal, meaning along the optical axis (LSA), or transversal (TSA), projected onto the paraxial image plane

That form of defect is best seen in a single lens magnifying glass. In the center of the lens, the image is best reproduced, but with looking more through the marginal regions, the image gets blurred. The reason is, that the focal point is rather a section on the axis as can be seen in figure 3.5. With increasing ray height from the optical axis, the ray is bent stronger, leading to an earlier intersection with the optical axis.

Coma Aberration

The name Coma directly indicates the shape of the resulting aberration, it produces a comet-like tail onto a point. This asymmetrical effect arises, when image points far off-axis get imaged, as shown in Figure 3.6.

A lens free of spherical aberration and coma is called an aplanat.

[1] don't get confused with the word paraxial. A paraxial ray is parallel to the optical axis, representing light from an object very far away. But the word paraxial is also used in the context of Gaussian optics, which uses the so-called paraxial approximation (3.4). In the context of general Snell's optics, the term paraxial region is also used when speaking about the region where $O(n^2) \approx 0$ in (3.2)

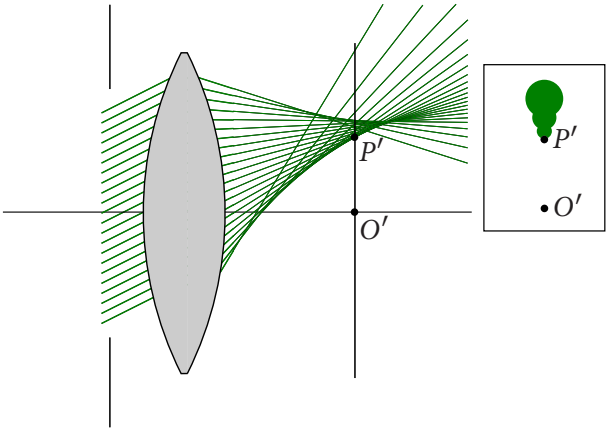


Figure 3.6: Rays from an off-axis point P with angle ψ to optical axis cause asymmetrical coma aberration. The longer the optical path, the farther away from focus in the image plane, hence the bigger the image point P'.

Astigmatism and Field Curvature

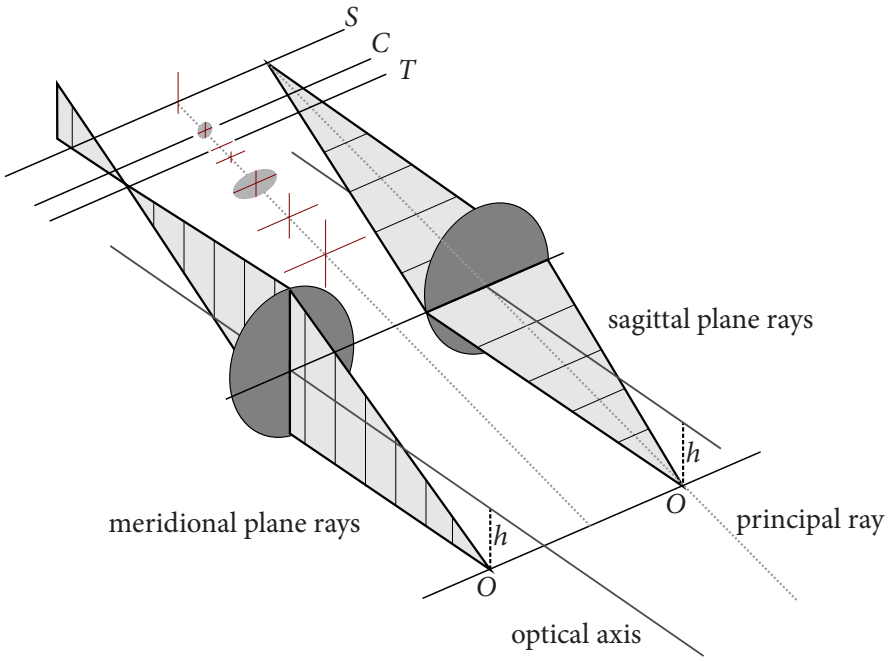


Figure 3.7: Identifiable variation of focus with raising ray height h. Between the T plane and the S plane there exists the circle of least confusion lying on plane C. The red crosses are for comparison of the particular ray fan width.

The spherical surface does hold some more surprises. The most abstract and unwanted one is Astigmatism. Imagine two orthogonal planes through the lens, both including the principal ray from origin O to image plane through the center of the lens. The meridional plane is the one also containing the optical axis, the other one is called sagittal plane. Now an off-axis point shifted along one axis in the object plane gets imaged as an ellipse onto the other side of a lens due to the asymmetric varying angle on the surface. For analysis we start rays lying in the before mentioned meridional and sagittal plane. The crux of the matter is the fact that those two fans have different focus points, like drafted in Figure 3.7. This a bit confusing image

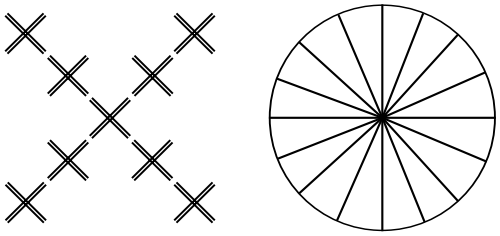


Figure 3.8: Both pattern consist of tangential and sagittal structures. Lenses that suffer astigmatism are not able to focus the whole pattern.

has to be clarified by examples. We have a circle, with the center lying on the optical axis. all points on the circle have the same focus plane for the rays in the tangential plane T, which means we can perfectly focus the whole circle by setting up the image plane exactly there. Lets add a line starting at the center of the circle and intersecting the ring. All points on this line have a different focus point behind T in the sagittal focus plane S. To put it in a nutshell, with an aplanat lens we are not able to exactly focus on structures with varying directions of edge extent. As the effect is depending on the ray height again, it becomes worse to the margins. A spoked wheel or an X-cross motif, like those in Figure 3.8 are the most common pattern to test for astigmatism. A lens free of all aberrations explained until now is called an anastigmat. Such a type now reveals a further defect. A curved imaging surface is not able to bring the object side to focus in a flat plane. It is more a paraboloid curved inwards to the lens in the positive case of field curvature. The surface is named Petzval surface after its discoverer Josef M. Petzval, who defined the theoretical base for the anastigmat lens. What we have not yet said is the obvious consequence, that a lens suffering astigmatism again, has two different curved surfaces of focus, one for the sagittal and one for the meridional plane. In figure 3.9 there is an example for the discrepancy between the different surfaces a lens can image on.

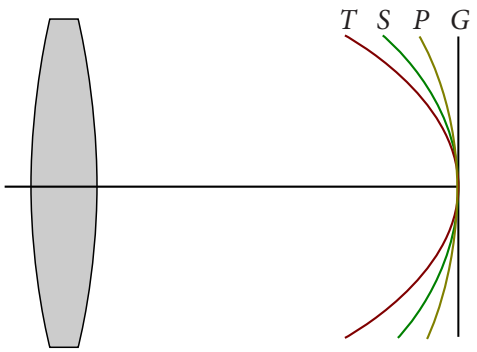
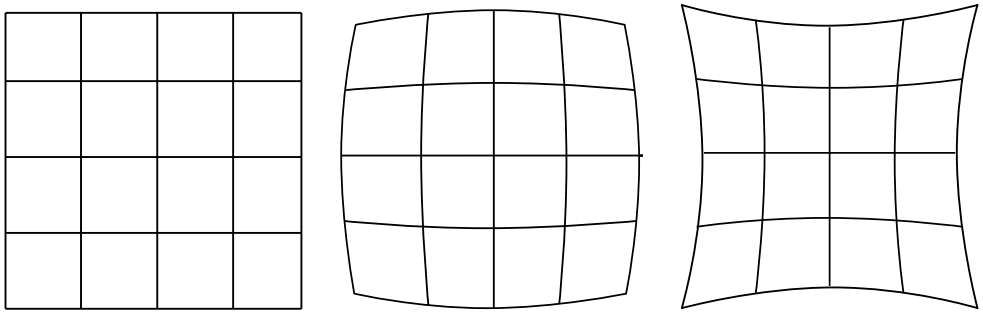


Figure 3.9: Field curvature is the consequence of spherical imaging elements. For aplanat lenses there exist two surfaces T and S, anastigmatic lenses image onto the Petzval surface P. G represents the Gaussian image plane.

Distortion

The last aberration in the group of monochromatic effects again leads back to the spherical surface of a lens. With the varying optical thickness for different ray heights we get a differently high amount of magnification. This can be evidenced by imaging a simple wire frame whose straight lines get resolved to more and more bent curves coming to the marginal region. In the center, we do not have to expect any distortion. With enlarging ray height, the image suffers either a reduction or boost in magnification,

according to the lens design and its quality. The term for inwards distortion is also pincushion distortion, where the magnification increases with ray height, the vice versa case is called barrel distortion.



(a) Undistorted wire frame (b) Barrel distortion (c) Pincushion distortion

Figure 3.10: Different types of distortion caused by varying magnification with lens thickness.

Chromatic Aberrations

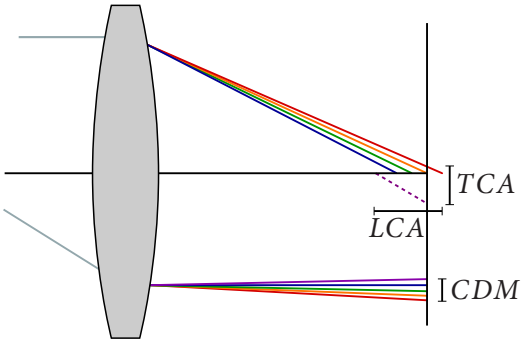


Figure 3.11: Dispersion causes image magnification varying with wavelength, called lateral color or chromatic difference of magnification (CDM). Longitudinal or transversal chromatic aberration (LCA/TCA) is the result of different focal distances for every wavelength with changing IOR.

The last defects to mention concern color imaging problems. As we already know from 2.1 the index of refraction in matter varies with the wavelength, which results in the effect of dispersion. Two different effects can then be distinguished which look a bit like wavelength dependant spherical aberration and coma. Figure 3.11 outlines the different types. Axial chromatic aberration either seen longitudinal or transversal is constant over the whole lens and forms the image in different focal planes. Lateral color from off-axis rays gets worse with increasing ray height, which means we can reduce the effect by cutting off marginal rays.

3.2.3 Aperture Inclusion

Optical systems for viewing are usually limited by stops and gates concerning irradiance transport and field of view [PPBS05, Smi07, Ray02]. Two main parts can be distinguished in duty. The one is the aperture stop, which is mainly responsible for irradiance control, by forming the cone of light which passes through the system from an axial point. Marginal rays, for example aberration critical ones become blocked there on the one hand. On the other, the overall irradiance can be also controlled by the generally adjustable

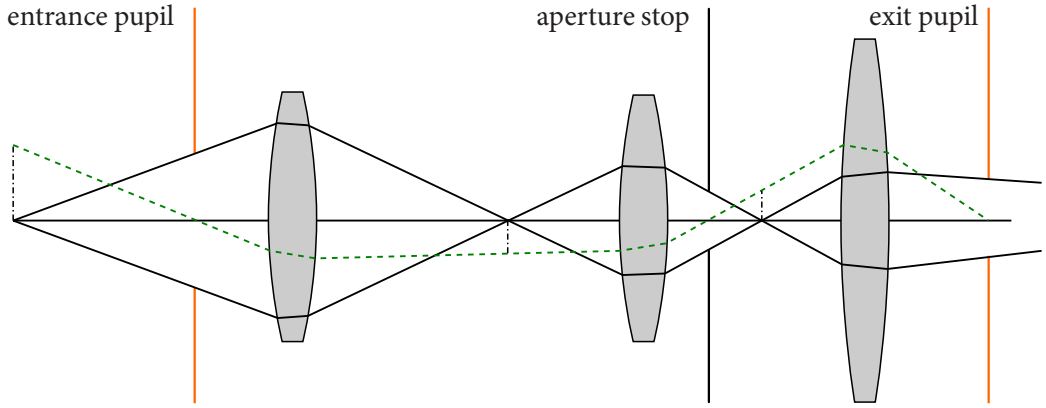


Figure 3.12: Adopted from [Smi07]. The (green,dashed) principal ray defines pupil locations, the axial ray fan determines the radii. The aperture stop is thus also a pupil.

diameter. The diaphragm construction consists of multiple blades that slide over each other to scale the gate in certain steps.

The field stop usually right in front of the light-sensitive layer limits the region of exposed film or sensor. Mapping the field stop through the lens to the object side results in the entrance window that also defines the angle of view according to the lens. The field stop has normally a fixed size to the appropriate format, while the aperture stop is variable in diameter by a diaphragm blade construction to form the aperture gate.

In relation to the aperture, the optician speaks of two pupils based on aperture size and position. They are a very useful information because the light forming character of a lens system can thus be reduced to two disks which can be seen as gates of contribution. Therefore we consider the ray traversal of some characteristic rays to make the construction clear. Figure 3.12 shows the ray tracing plot of a three element example system. The two main rules for pupil construction are [Smi07]:

1. *Principal ray intersections with the optical axis yield the pupil locations.*
2. *The cone of rays starting on the object side on-axis, passing the aperture yields the pupil radius when viewing at the cone diameter in the pupil locations.*

In the next step, we need to consider the aperture as controlling mechanism in detail. The photographic f-number N can be derived by

$$N = \frac{f}{d_{entr}}$$

normally written as f/N with focal length f and entrance pupil diameter d_{entr}. For a multi element device, the formula appropriately extends to:

$$N = \frac{f_{eff}}{d_{entr}} \tag{3.4}$$

where f_{eff} is now the effective focal length resulting from the combination of the lenses, and d_{entr} is the diameter of the entrance pupil as representative on the object side. The effective focal length can be evaluated by paraxial optics tools, not further described here, because this number is already denoted for given lens designs. For detailed studies, be referred to [Ray02, Smi07, Smi05].

As a consequence of the adjustable aperture diameter, a lens is able to have multiple f-numbers it can be set to. This allows for controlling the effective irradiance passing the lens and thus the reaching intensity on the sensor. The not so far introduced camera exposure value of the resulting image varies by changing the aperture diameter. There exists a normed table for photographic purposes which is not further mentioned

stop value	0	+1	+2	+3	+4	+5	+6	+7	+8	+9
f-number	$f/0.5$	$f/0.7$	$f/1.0$	$f/1.4$	$f/2.0$	$f/2.8$	$f/4.0$	$f/5.6$	$f/8.0$	$f/11.0$
aperture area	100%	50%	25%	12.5%	6.25%	…				

Table 3.1: Standard full stop scale with given stop values and relative aperture area for a hypothetical $f/0.5$ lens.

here because of additional dependence on the exposure time [Ray02]. For our setting we are satisfied with knowledge about the f-scale, that evolved for certain reasons. The linear variation of aperture radius results in a quadratic change of illumination. That can be easily shown by the circle area equation

$$A_{ap} = \pi r^2, \tag{3.5}$$

which means by halving the radius, the passing illumination gets quartered. A consequence is the root of two factor of ≈ 1.414 in the standard f-number scale, specified in Figure 3.1. An impression on the change of intensity is also given. One step up on the standard full-stop scale means halving the irradiance through the aperture. In this context, the photographer speaks about the *speed* of a lens, because obviously, the larger the maximum aperture, the less time is needed for exposure. Decreasing the diameter by one full step is called *stopping up*, whereas *stopping down* means opening the diaphragm. As we already know from the previous sections, aperture size does not only influence irradiance. Marginal rays can be eliminated by a high f-number, which minimizes aberrations. Thus we cannot say that the aperture is solely responsible for irradiance control.

3.2.4 Glass Science

Glass is the main resource for optical engineers, besides plastic and crystalline materials. In 1884 Otto Schott founded the *Glastechnische Laboratorium Schott & Genossen*, with valuable support of Ernst Abbe [Ray02, Per07]. They soon released a catalog with numerous types of glass with specific optical properties mainly concerning the refractive index and degree of dispersion. Today there are many other glass attributes which must be taken into account for special purposes, like behaviour in different temperatures, but they can be neglected for this topic. An excerpt from the actual catalog from the year 2009 gives a short example in Table 3.2.

code	η_d	η_e	V_d	V_e	B1	B2	B3	C1	C2	C3
N-SF1	1.717	1.723	29.62	29.39	1.609	0.238	1.515	0.01197	0.05906	135.522
N-SK11	1.564	1.566	60.80	60.55	1.179	0.223	0.936	0.006803	0.02199	101.513

Table 3.2: Schott catalog entries for flint glass SF1 and crown glass SK11 [Sch09]. The C^* values have the unit $[\mu m^2]$.

Due to the dispersive character of glass, the refractive index has to be regarded in different wavelengths. The Abbe V-number is a common description for the dispersion characterization in the visible spectrum [Smi07].

$$V = \frac{\eta_d - 1}{\eta_F - \eta_C} \tag{3.6}$$

The higher V, the lower is the dispersive behaviour of the material. The subscripts of the refraction indices specify the according wavelengths by letters of the Fraunhofer lines, in particular the F-line at 486.1nm, the d-line at 587.6nm and the C-line at 656.3nm. To resolve a certain IOR for any given wavelength hereby is not very useful, as we wish to use the hole visible spectrum. Indeed, dispersion is not linear with wavelength

and there exist only empirical relationships described for example by the three-term Sellmaier equation which is used by Schott definitions today:

$$\eta(\lambda[\mu m]) = 1 + \frac{B_1\lambda^2}{\lambda^2 - C_1} + \frac{B_2\lambda^2}{\lambda^2 - C_2} + \frac{B_3\lambda^2}{\lambda^2 - C_3} \tag{3.7}$$

The different types of glass and their combination play a main role in lens design and aberration correction.

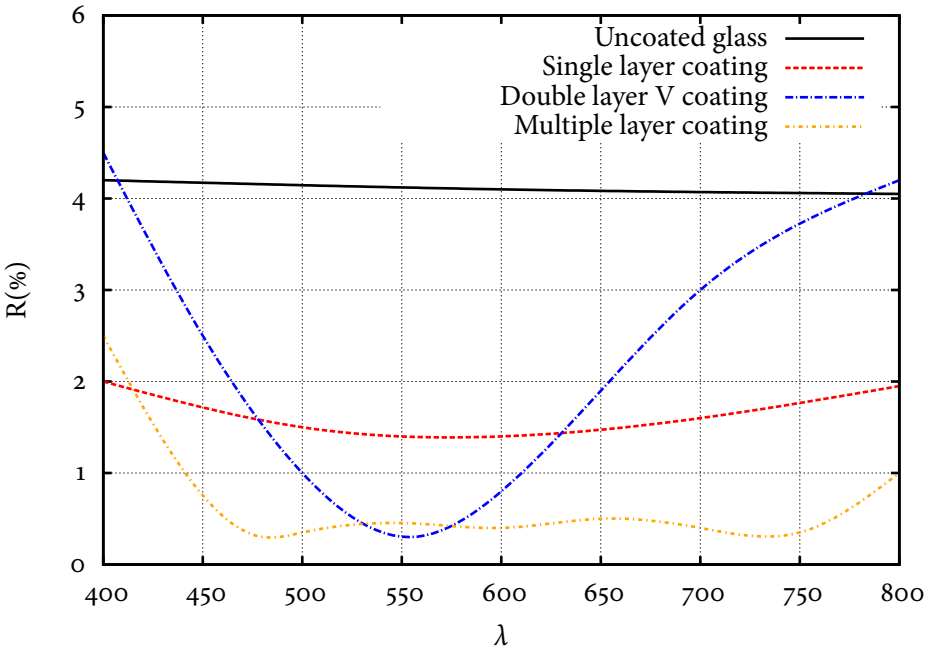


Figure 3.13: The comparison of different coating setups in degree of reflection shows that we are able to regain a high percentage of reflected light. The quality and range depends on field of use and costs. Data taken from [Smi07].

3.2.5 Lens Design Data

Today, lens designers have access to a plethora of standard designs to start with new drafts. A very popular representative is the Tessar design originally patented 1902 by Zeiss. With their names, certain attributes are predetermined and well understood.

To become more familiar with descriptions like Figure 3.14, here some additional words. Every line in such a table represents one surface, from the beginning we start with the leftmost surface. The first number expresses the curvature radius of the surface, its sign behaves within the rules of optics stated before. The second value describes the material thickness along the optical axis. The material name is the code originating from the Schott glass catalog, one of the main references for lens designers, which includes data about refractive properties. The columns about index of refraction (IOR) and Abbe V-number (V-no) are only mentioned for fair orientation. The last column represents the semi-diameter of the element. In line 1 and 2 of Figure 3.14 two different types of glass follow directly behind each other. That denotes a group or compound, which is glued together with special cement. Last to mention are the lines 3 and 4 which look like a mistake, because at first sight, an airy surface follows an airy surface. That indicates the presence of an aperture stop inside the system. To describe the whole system from Figure 3.14 in one sentence one could say: "Here we have a reversed[2] Tessar lens design in three groups with four elements, created for

[2] in the original Zeiss patent, the compound was at the film-side end of the system

rad	thick	matl	IOR	V-no	sap
42.970	9.8	LAK9	1.691	54.7	19.2
-115.33	2.1	LLF7	1.549	45.4	19.2
306.840	4.16	air			19.2
	4.0	air			15.0
-59.060	1.870	SF7	1.640	34.6	17.3
40.930	10.640	air			17.3
183.920	7.050	LAK9	1.691	54.7	16.5
-48.910	79.831	air			16.5

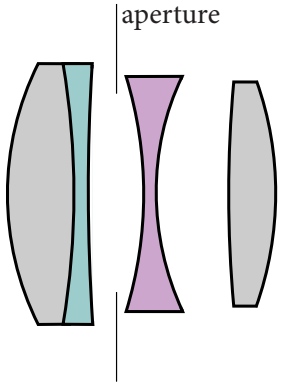


Figure 3.14: Tessar Design by Brendel (USP 2854889) [Smi05], $f/2.8$, 100mm effective focal length (EFL).

maximum speed $f/2.8$." So much for optical engineers' talk.

The cone of illumination does not neccessarily expose the whole image sensor, because every lens has a maximum geometric image height it is able to produce. For the Tessar example we have an imaging height of approximately 52mm at 100mm focal length. If we want to use it in combination with a 36x24mm image sensor, the diagonal to the corners must be smaller than $\sqrt{18^2+12^2} \approx 21.63$mm. Thus we are allowed to reduce the focal length by the factor of at most $frac5221.63 \approx 2.4$. The scaling of focal length for a given lens design is relatively simple. We are allowed to multiply every single value of extent by a desired factor to reduce or enlarge a design [Smi05].

3.2.6 Material Coatings

It is requested that an optical system ideally delivers all incident light to the image plane. In reality, this is prevented by the formerly introduced Fresnel Equations. For light at normal incidence (2.5) degenerates to

$$T = \frac{\eta'}{\eta}\left(\frac{2\eta}{\eta+\eta'}\right)^2 \tag{3.8}$$

which shows that full intensity transportation is physically impossible. Note also, that polarization has no impact on the result because all $\cos\theta$ terms diminish. For uncoated glass with $\eta = 1.6$ (3.8) delivers approximately 95% of transmission or 5% reflection. When recalling the Tessar setup in Figure 3.14 it becomes even more clear, that lacking transmission can become a serious problem. For this design with uncoated glass, the transmission falls back to around 70%.

The basic idea behind lens coating is, to increase the part of refracted light by thin layers with special intermediate refraction indices. In order not to step too far into the topic of thin-film interference and complex arithmetics, it will suffice to show the effect of coatings in Figure 3.13. For more details on the used substances, and math behind it, please have a look at [Ray02, Smi07].

3.2.7 Secondary Effects

This short subsection gives a brief description of some secondary effects, that especially arise, as soon as a lens design is used beyond its limits or is not sufficiently refined with coatings.

Flare Effects

Light in a lens mount is not only refracted towards the sensor. The Fresnel Equations in (2.5) show, that reflection is an inherent effect, that can be reduced by coating but it is still there and can be made visible in several situations. A typical flare effect is oblique incoming light, that gets reflected at a surface and then reaches the image plane [Smi07]. An example is given in Figure 3.15.

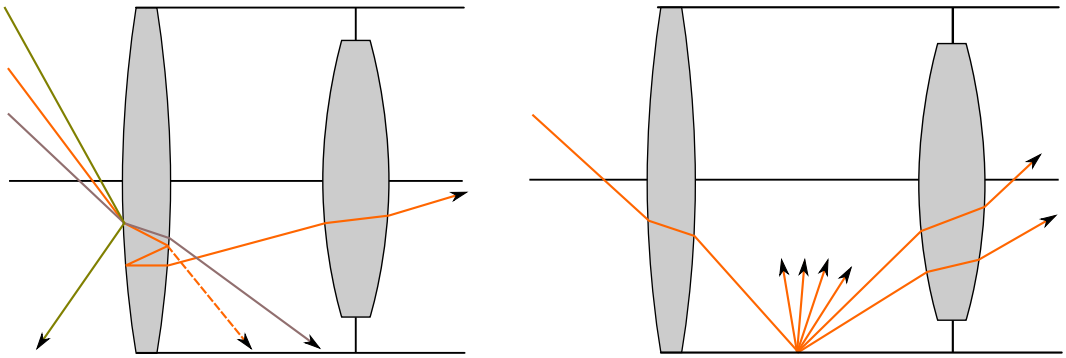


Figure 3.15: Secondary reflections inside the lens barrel can arise due to extreme angles of incidence, which cause internal reflections inside the elements, as seen on the left. The right schematic shows, that stray light occurs when the barrel has a certain reflectance.

Typical countermeasures are for example lens hoods, to block oblique rays outside the field of view before they enter the lens system. Professional camera bodies and lens mounts are furthermore plated from the inside with a black, highly diffuse sort of small light barriers, called baffles, which block incoming light and scatter it back towards the entrance pupil [Smi07]. But as mentioned, under extreme conditions, any optical device reveals these effects. Figure 3.17 gives a nice example. These flare effects can expand to whole ghost images of bright parts, that arise, when when a highly reflective glass surface was not coated.

Vignetting

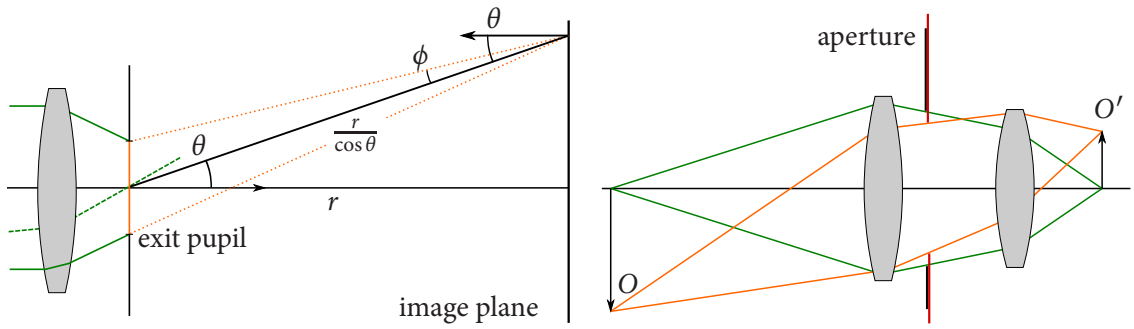
The effect of non-equal illumination of the imaging area is called vignetting and basically consists of three independent subtypes [Ray02]. The first is also called natural vignetting and is quickly explained by recalling Figure 2.7b. The incoming irradiance from a lens onto the sensor is subject to exactly the same effect assuming the exit pupil to be an area light source. Thus we must introduce a further compensation factor, because the angle from the exit pupil plane normal to the direction has also to be accounted for. This angle is almost constant for large distances between exit pupil and image plane, respectively small exit pupils, where $\phi \approx 0$. It follows the *cosine-forth-law* of illumination which holds for the just explained circumstances:

$$H_\theta = H_0 \cos^4 \theta \tag{3.9}$$

where H_0 is the radiant exitance along the normal, and θ the angle from the illumination line to the normal of either planes. See Figure 3.16a for a combination of Figure 2.7b and the recently introduced extension. For a correct compensation of irradiance for influential sizes of ϕ be referred to [Smi07].

Optical vignetting is a consequence that off-axis rays have another *effective aperture*, because not all elements of the system can accept the formed cone of illumination [Smi07]. Figure 3.16b drafts two different cases. The reduced aperture area of oblique rays causes an appropriate decrease of irradiance at the image plane.

The last type of vignetting, also called mechanical vignetting, is due to shadowing by lens hoods or lens bodies of wrong dimension that influence the field of view [Ray02].



(a) Natural vignetting (b) Optical vignetting

Figure 3.16: Natural vignetting can be directly derived from previously explained irradiance issues. Optical vignetting is caused by the lens design. The aperture cannot be completely imaged because of the limited diameter of the lens in front and behind it.

Figure 3.17: Picture taken with a Canon 5D MKII and Canon L Series lens. The two different colors of the lens flares reveal a bit about the different used coatings. The mid enlargement of the house wall shows another alerting effect. The contrast of the wall looks considerably reduced, which can happen due to stray light in the lens barrel or veiling glare. The latter mentioned is a lens flare that spread over the whole image area.

Figure 3.18: This panoramic fisheye picture of the *Ulmer Muensterplatz*, taken with a Peleng 8mm circular fisheye lens on a Canon 5D MKII at $f/16$ shows various imaging defects, mentioned in this chapter. What first catches the eye is the deliberate distortion which is needed to reach the 180° field of view. In the upper left corner there is a bright ghost of the sun, formed behind the aperture stop, thus it has the angled diaphragm shape. Explicit chromatic aberration on the edge of the ghost is visible, also along the whole image border. In the lower right corner we see glaring stray light which emerges from reflection of the sunlight inside the body.

4 Photo-realistic Image Synthesis

The history and development of ray tracing in computer graphics as tool for photo-realistic rendering is a vast field with many topics accrued over the last 40 years. Most basic ideas outlined here can be deepened with lecture of [Shi05, Wat02]. On the following pages describe briefly the most important approaches and algorithms used for the implementation of CHROMA, the rendering software that evolved during the composition of this thesis.

4.1 Image Generation by Ray Casting

4.1.1 Rasterization Principle

A basic idea of ray tracing follows the ancient idea that the perception of light is caused by rays leaving the eye and hitting the environment. The scene to be imaged is described in geometric primitives which allow us to analytically evaluate intersection points p for given rays $R(o, \vec{d})$, with origin o and direction $\vec{d}$. This is the contradictory approach of image generation in contrast to graphics cards algorithms that rasterize every triangle into an image and Z-Buffer [Shi05]. By ray tracing, we sample all given scene primitives by rays, to compute a color for every pixel.

Ray-Plane Intersection

The most basic intersection test is the ray-plane intersection [Shi05]. A plane H is defined by a point in space p_H and a normal $\vec{n_H}$ with $|\vec{n_H}| = 1$ for orientation. The intersection point can be analytically solved with the implicit equation for the plane

$$(p - p_H) \cdot \vec{n_H} = 0 \tag{4.1}$$

The ray equation

$$p = o + t\vec{d} \tag{4.2}$$

inserted to (4.1) resolved for t is

$$t = \frac{(p - o) \cdot \vec{n_H}}{\vec{d} \cdot \vec{n_H}} \tag{4.3}$$

with caution of the possible cases $\vec{d} \cdot \vec{n_H} = 0$ and $\frac{0}{0}$.

Ray-Triangle Intersection

The simplest description of a finite area is the triangle T with three vertices v_0, v_1, v_2 and a normal $\vec{n_T}$. The ray-plane intersection can be applied directly. The test whether the plane hit point lies inside the triangle bounds can be done by barycentric coordinates. They describe every point inside the triangle the following way:

$$p = \alpha v_0 + \beta(v_1 - v_0) + \gamma(v_2 - v_0), \quad \alpha + \beta + \gamma = 1 \quad \wedge \quad \alpha, \beta, \gamma > 0 \tag{4.4}$$

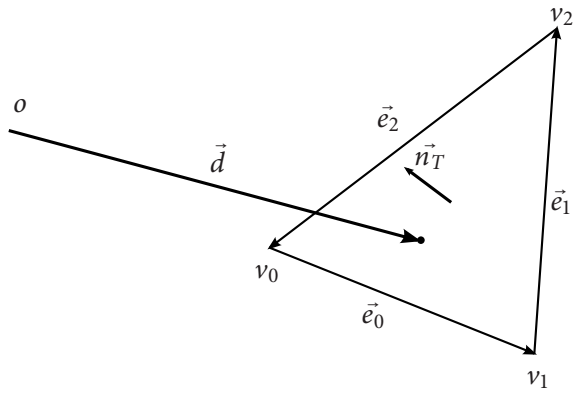


Figure 4.1: A Ray-triangle intersection test is usually based on the idea of checking, whether the point lies inside the bounds of the edges.

A well established, simple, and fast method, is for example the "Moeller-Trumbore" intersection test [MT97], given as reference for further reading in this place.

Ray-Sphere Intersection

The evaluation of a ray-sphere intersection can now obviously deliver two hit-points which makes the computation a bit more difficult. For a point p on a sphere with radius r and center c it applies:

$$(p-c)\cdot(p-c) = r^2 \tag{4.5}$$

Including equation (4.2) yields

$$(o+t\vec{d}-c)\cdot(o+t\vec{d}-c) - r^2 = 0$$

rearranging to the quadratic equation

$$(\vec{d}\cdot\vec{d})t^2 + 2\vec{d}(o-c)t + (o-c)\cdot(o-c) - r^2 = 0$$

solves for

$$t_{1,2} = \frac{-d(o-c) \pm \sqrt{(d\cdot(o-c))^2 - (d\cdot d)((o-c)(o-c)-r^2)}}{d\cdot d} \tag{4.6}$$

Now we can check both results, which point is first along the intersection ray and return the closer one not negative. If we started our ray inside the sphere, we also have to keep in mind that we have to reject a negative result for t. For rays passing the sphere, the discriminant becomes smaller than zero, which means we only have to evaluate the root operation once we hit the sphere surely.

Ray-Dome Intersection

A spherical cap, or dome, is part of a sphere, defined by the latter and an additional cap height h or alternatively a base radius r_{base} [BSMM05]. It is obvious to start the test with a normal sphere intersection. After receiving the values $t_{1,2}$ we have to restrict the pool of solutions to the angle given by the sphere cap. Therefore a vector c_{dome} to the center of the dome is defined for the purpose of orientation. Algorithm 1 outlines the needed operations.

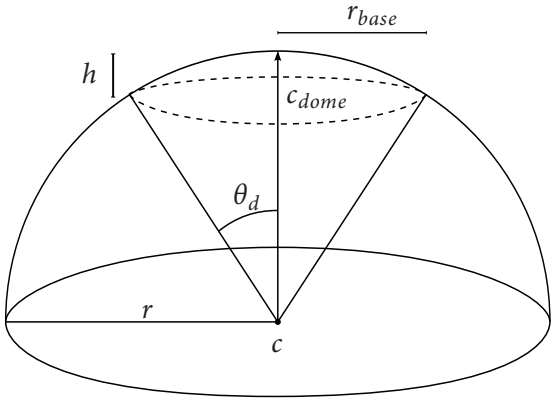


Figure 4.2: Spherical dome description

Algorithm 1: Ray-Dome Intersection

```
   Data: spherical dome, ray
   Result: first hit point on the dome surface
 1 t1,t2 = intersect(ray,sphere(dome.curveRadius,dome.center));
 2 if t1>t2 then
 3 |   swap(t1,t2);
 4 end
 5 float t=t1;
 6 domeTest:
 7 if t>0 then
 8 |   Vector3 hitpoint = ray.at(t);
 9 |   Vector3 hitnormal = (hitpoint - dome.center) / dome.curveRadius ;
10 |   float cosAngle = hitnormal · dome.vectToTop;
11 |   if cosAngle>dome.cosSemiAngle then
12 |   |   return hitpoint;
13 |   end
14 |   t=t2;
15 |   goto domeTest;
16 end
```

Further useful formulae are:

$$A_{dome} = 2\pi hr \tag{4.7}$$

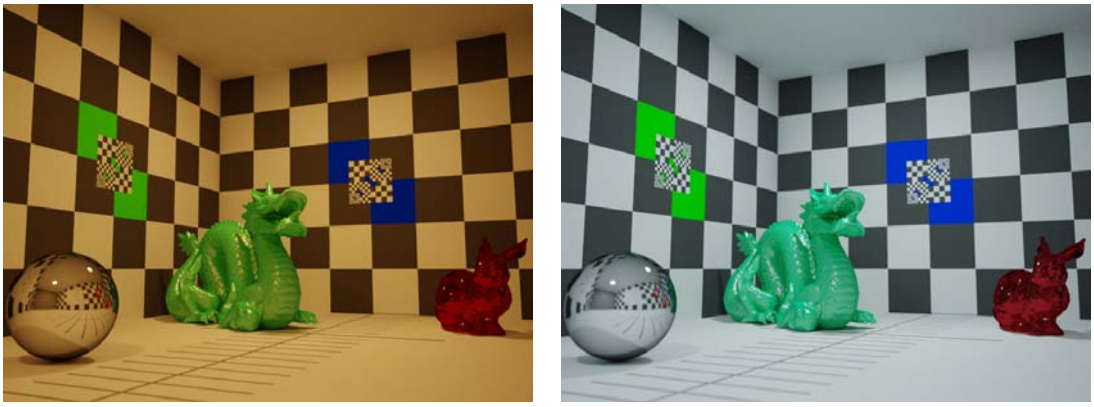
$$h = r - \sqrt{r^2 - r_{base}^2} \tag{4.8}$$

$$\theta_d = \cos^{-1}(1 - \tfrac{h}{r}) \tag{4.9}$$

Sub-linear Scene Primitive Traversal

With increasing number of primitives in a scene we come to the problem that the number of tests also increases linearly, which becomes terribly inefficient when speaking about thousands of triangles for example. Therefore several acceleration structures exist with different concerns and features. Just to name two popular representatives, there are the *kD-Tree* and the *BVH-Tree*. I used the latter one in my implementation. Further details on that subject and classification can be found in standard graphics literature [Shi05, Wat02].

4.1.2 Spectral Rendering



(a) Test scene rendered with CIE A Illuminant data (left) and D65 light source model for comparison (right).

(b) Clearly visible dispersion due to deliberate multiple total internal reflection in a dielectric. Rendered with an environment map, a simple and efficient way to approximate light sources.

Figure 4.3: Spectral rendering examples, created with an early revision of CHROMA.

To receive a coloured image consisting of RGB values per pixel, normally RGB input values are also used for material definition. An RGB value can be seen as a triple of three specific wavelengths which will be accounted for. Indeed, all rendering procedures also apply to a one-dimensional value, which reduces the setting to monochromatic *spectral rendering*. Now at that point we could replace the fixed wavelength per ray by a random wavelength and apply this one to a ray used for the image illumination computation. The result is a spectral radiance value, which can be mapped to a color vector again in the accumulation step.

Spectral ray tracing allows us to adopt all previously mentioned physical laws to the rendering environment. This includes wavelength dependent dispersion from Section 2.1.3, diffraction from Section 2.2, and the usage of spectral power distribution data for light sources as shown in Section 2.3.

The main problem in spectral rendering, is the fact that one needs huge data sets for BRDF representation of every material used in the scene. They are acquired for example by gonioreflectometer measurements for every wavelength. For materials with a certain structure or color variations, the process must be done for the whole surface at some discretization.

To limit the effort of both spectral reflectance data handling and to allow RGB texturing, we use the conversion method for RGB reflectance values by Brian Smits [Smi99]. This approach allows to estimate the equivalent spectral reflectance for a given wavelength and a RGB triple. The algorithm delivers sufficient

results under consideration of the fact that an RGB value has an infinite number of spectra as representatives which are all perceived the same. This circumstance is called metamerism, already introduced in Section 2.4. A further problem is that a spectrum is usually a continuous function of arbitrary distribution. The algorithm discretizes the spectrum to a number of piecewise constant sub-spectra together with a weight vector; we could also say we now use n-dimensional color vectors. Because the conversion from a spectral radiance result back to RGB over the XYZ color space is linear and well understood, it offers the possibility to render the whole image in XYZ and convert it back in a post process. For example white balance operations due to white point definitions can be handled much better, because they are implicitly included in the conversion matrix. Because spectral rendering does not belong to the direct field of concern of this thesis, I refer to standard literature discussing Colorimetry for conversion to RGB [Per07, Shi05] and to the paper by Smits for further reading and algorithms covering the conversion to spectral reflectance [Smi99].

The rendered images in Figure 4.3 show results of the implementation of both algorithms by Smits and the usage of different spectral power distributions from Figure 2.8.

4.2 Global Illumination and the Monte Carlo Method

The generation of images by ray tracing is basically done by evaluating the paths emanating from light sources, reflected on scene surfaces, and thus arriving on the sensor. The term global illumination in that case means, that we account for as many light interactions as we can or better, as we *need*. In order to solve the light transport integral over a surface, numerical evaluation by the Monte Carlo Method. Eric Veach introduced a very good compendium about this field specialized in light transport computation [Vea97]. The most important terms and approaches are pesented in the following sections.

4.2.1 Probability Theory Basics

This is just a very brief roundup on that vast field, to understand the later formulae concerning the Monte Carlo Method.

Let X be a discrete random variable and x be realisations of X with the following definition of probability: In a probabilistic experiment where the specific realisation x occurs in n cases of N total events, the probability of the realization x is

$$P(x) = n/N. \tag{4.10}$$

The cumulative distribution function (CDF) of a random variable X returns the probability that a realisation x is smaller than a given constant c

$$F_X(c) = P(x \leq c). \tag{4.11}$$

In contrast, for a continuous random variable Ξ, we are not able to give a probability for one realisation ξ, instead for an interval $[a, b] \ni \xi$ applies

$$P(a \leq \xi \leq b) = \int_a^b p(t) \quad dt \tag{4.12}$$

where $p(t)$ is the probability density function (PDF) with the following relation to the CDF:

$$F_\Xi(c) = P(\xi \leq c) = P(-\infty \leq \xi \leq c) = \int_{-\infty}^{c} p(t)dt. \tag{4.13}$$

The expected value of a continuous/discrete random variable Ξ/X can be derived by

$$E(\Xi) = \int_{-\infty}^{\infty} p(t)\,t\,dt, \tag{4.14}$$

$$E(X) = \sum_k x_k P(x = x_k). \tag{4.15}$$

The quadratic deviation or variance is defined as

$$V(\Xi) = E((\Xi - E(x))^2), V(X) = E((X - E(X))^2).$$

Some further rules can be acquired to make analysis of variance and expected value easier:

$$E(a\Xi) = a\,E(\Xi) \qquad V(a\,\Xi) = a^2\,V(\Xi)$$

$$E\left(\sum_{i=1}^{N} \Xi_i\right) = \sum_{i=1}^{N} E(\Xi_i) \quad V\left(\sum_{i=1}^{N} \Xi_i\right) = \sum_{i=1}^{N} V(\Xi_i)$$

A random variable could be part of a function f which allows the straight forward substitution

$$\Psi = f(\Xi). \tag{4.16}$$

Thus, all recently mentioned formulae also apply for functions.

4.2.2 Monte Carlo Integration

With given basics, we would like to estimate the integral of a function $f(x)$. For that, we evaluate it at discrete random positions with realisations x_i of an independent and identically distributed (i.i.d.) random variable X and sum the results up with respect to the probability of the sample, which leads to an estimator S of the function f:

$$S = \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{p(x_i)} \tag{4.17}$$

If the realisations are uniformly distributed on $[0,1]$ the estimator degenerates to the special case

$$S = \frac{1}{N} \sum_{i=1}^{N} f(x_i) \tag{4.18}$$

The quality of the estimator S now depends on some factors. First, it is obvious that we only sample the region of interest, which follows that $x \in [a,b]$. Furthermore our solution depends on the function and the distribution of our random numbers. The expected value of our estimator proves correctness by using (4.15), and

$$E(S) = E\left(\frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{p(x_i)}\right) = \int_a^b f(x)\,dx. \tag{4.19}$$

The operation is valid because the estimator itself can be seen as a random variable again. It must apply that $p(x) \neq 0$ over the sampling domain, which can be understood as the probability to miss an important part. It is important to note, that it is not correct to say that the result of the estimator is the same than the result of the integral, but the consideration of variance shows an important connection:

$$V(S) = V\left(\frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{p(x_i)}\right) = \frac{1}{N^2} \sum_{i=1}^{N} V\left(\frac{f(x_i)}{p(x_i)}\right) = \frac{1}{N} V\left(\frac{f(x)}{p(x)}\right) \tag{4.20}$$

by using the above explained rules. Now the result is, that with increasing the number of samples, the variance of the estimator decreases to zero. The root mean square (RMS) error or standard deviation $\sigma(S)$ is simply

$$\sigma(S) = \sqrt{V(S)} = \frac{1}{\sqrt{N}}\sigma X \tag{4.21}$$

where X again is the random variable included in the estimator S. The convergence rate of a Monte Carlo estimator is thus $O(N^{-\frac{1}{2}})$.

Another more abstract point of view offers the *law of large numbers* that also holds for an estimator with infinite variance:

$$prob\left\{E(x) = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} x_i\right\} = 1 \tag{4.22}$$

But in general we do not have time to sample an infinite number of positions, so we have to be satisfied with the conclusion

$$\int_{x\in D} f(x)\,dx \approx \frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{p(x_i)} \tag{4.23}$$

where now D represents an arbitrary domain of events.

4.2.3 Inversion Method

To gain samples x from a given arbitrary distribution we can apply the inversion method and construct F^{-1} [BSMM05]. Provided that, F is a strictly monotonic continuous function with $0 \le F(x) \le 1$, there exists the inverse function F^{-1} with

$$\forall u \in [0,1]_{uniform} : F^{-1}(u) = x \in D \tag{4.24}$$

where x suffices the probability distribution of F and the according density p.

In the discrete case, the technique is based on an intuitive chain of operations. Given a set of random events $x_i \in D$ with probabilities $P(x_i)$, we can construct the cumulative distribution by hand:

$$F(c) = \sum_{x_i \le c} P(x_i) \tag{4.25}$$

Under the above stated properties of F we are now able to sample the inverse CDF by inserting a uniformly distributed value u and return the event x_i being valid for this interval of P.

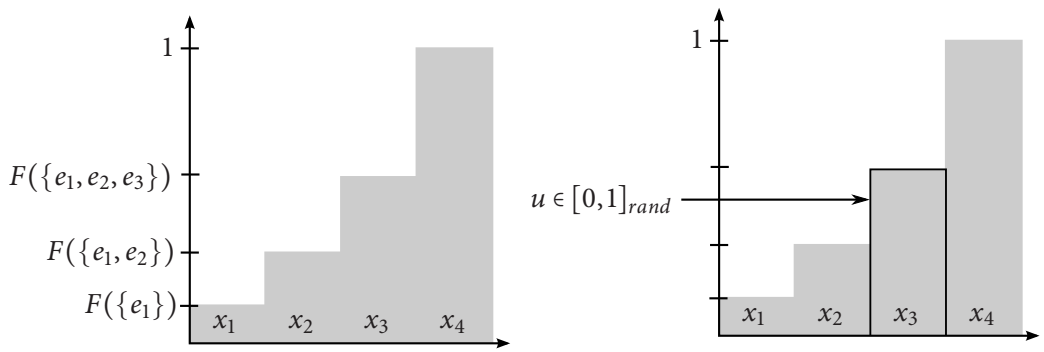


Figure 4.4: An example for creating random events x_i by a constructed cumulative distribution function and using its inverse.

4.2.4 The Rendering Equation

In 1986 most of today's rendering techniques by ray tracing were already known and were finally formulated in general by James Kajiya in the rendering equation. This is the basic formula to start with, when introducing physically based global illumination. Now, the radiometry formulae from section 2.3 will be of use.

$$L(x,\omega_o) = L_e(x,\omega_o) + \int_{S^2_-} \rho(\omega_i,x,\omega_o)L(y,\omega_i)\cos\theta_i d\omega_i \tag{4.26}$$

This reads: *Light from a surface point x in direction ω_o is emitted radiance L_e at x in that direction plus incident radiance L from all directions ω_i over the hemisphere S^2_- reflected into ω_o by the bidirectional reflectance distribution function (BRDF) ρ.* The equation should look familiar, when recalling (2.21) and following examples in that section. The evaluation of this equation offers two main challenges. First, the result depends recursively on other radiances, and second, we need to come to another more comprehensive unit, like irradiance or intensity, to be able to compute an image at all, because radiance is a direction dependent unit.

This equation obviously applies only for vacuum transport, because scattering in media was neglected.

Operator Formulation of Light Transport

Before we come to deeper details of path tracing, let us have a look on an alternative way of writing down (4.26). For the ease of light transport formulation, the operator notation is very beneficial for describing the transport equation with otherwise huge integral terms:

$$L = L_e + T_\rho L \tag{4.27}$$

The transport operator T_ρ stands for the whole transport of indirect radiance L which started at a light source at the beginning of the path. Which interactions the path had is also of high importance and will further be written by Heckbert's regular expression notation for light paths [Vea97]. The general form is $L|(S|D)*|E$, where L stands for a light source location, S for a specular scattering event, D resembles a diffuse scattering incident, and E means the path is located on the image recording sensor. I extend this formulation by a further letter T for transmission, to separate dielectric locations from the rest of the other events.

The Pixel Integral

As we would like to evaluate the incoming intensity on a pixel with given area A, we have to integrate over this area and collect all incoming radiance over the hemisphere. This results primarily in irradiance per location x,

$$H(x) = \int_{S^{2-}} L(x,\omega)\cos\theta d\omega, \tag{4.28}$$

then it follows that

$$I_{pixel(x,y)} = \int_A H(x)dx. \tag{4.29}$$

Path Space Sampling

As soon as we want to adapt the presented stochastic theories, we come to the imagination border, what are the events in a light transport simulation?. We have to sample different stages of a light path, for example

by choosing an emission direction. Sampling a direction into a hemisphere looks like

$$\begin{aligned} u_1, u_2 &\in [0,1]_{rand} \\ \phi &= 2\pi u_1 \\ \theta &= \cos^{-1} u_2 \qquad p((\phi,\theta)) = \frac{1}{2\pi} \end{aligned}$$

where ϕ is the angle on the equator and θ describes the angle to the pole cap [Duto3]. The transformation to a z-up Cartesian coordinate system can be done by

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \cos\theta \sin\phi \\ \sin\theta \sin\phi \\ \cos\theta \end{pmatrix} \tag{4.30}$$

In fact we only constructed a uniformly distributed random point on a unit hemisphere and used it as a direction. What we can do now, is solving for irradiance at a location x, by sampling into the unit hemisphere oriented along the surface normal and looking for radiance coming from sampled directions [Duto3].

$$H \approx \frac{1}{N} \sum_{i=1}^{N} L(h(x,(\phi,\theta)), -(\phi,\theta)) \cos\theta \tag{4.31}$$

The function $h(x,\omega)$ is called ray shooting function which stands for more or less the whole concept of ray tracing noted before. The second step in (4.29) is to evaluate the area integral over the pixel. It is exactly the same procedure, this time not for directions but location samples. We can choose random points on the pixel , evaluate H there and again compute the solution by (4.18). Here it becomes clear, that we simply compute the average over the surface irradiance during integration which could be weighted by an optional filter kernel. A uniformly distributed sample on a rectangular surface, for example a pixel, can simply be gained by two random numbers scaled to the appropriate extent. The PDF is uniformly $\frac{1}{A_{surface}}$, which applies to every uniformly generated sample on a surface. In the following, the inverse probability is also called *path weight*. This originates from the idea, that during path construction, all weights introduced by sampling can be multiplied to yield a combined probability or importance. Generally, all factors considering the received radiance can be inserted here. Finally the whole transport operator T_ρ is one large path weight which consists of reflection events, direction sampling and all other radiance influencing multipliers.

4.2.5 Material Representation

Before we finally come to light transport simulation, we need a description of how light can be transported by surfaces at all and how it can be integrated into the transport operator T_ρ.

With enough effort, we are able to simulate light transport through and over almost every desired surface. To cover this section in detail, be referred to modern physically based rendering literature [PH04]. The BRDF already introduced in 2.3 can be described in many different analytic ways. The model included in CHROMA is the widespread Phong reflection model, which includes a diffuse and a specular term. For further discussion, especially on the reciprocal and energy conserving version, [LW94] is recommended. Just to mention the abstract part of sampling for further comprehension, i give a short example: We hit a Phong surface by a path with actual weight $w(x)$ at the last hit-point x. The weight did not change, because we assumed vacuum transport. The Phong surface can be seen as a two-layer material, consisting of a glossy coating over a diffuse surface. Which layer we evaluate, can be decided by the inversion method, the probability $P(e_{Phong})$ for both events can for example be chosen by the length of each colour vector, with

the sum normalized to one. In a nutshell, the construction of a continuing ray at hit-point y makes the following changes to the path-weight w:

$$w(y) = w(x)\frac{1}{P(e_{Phong})}\rho(x \to y, y, \omega_o)\frac{1}{P(\omega_o)}$$

where e_{Phong} is the sampled Phong event, and ω_o the new direction. The contribution of wavelength and direction is included in ρ.

Glass and water are important surface materials to produce images with exciting light caustics. In addition, for a realistic calculation of rays passing a lens, the laws of geometric optics concerning refraction are inevitable. Thus the formulae from section 2.1 directly apply to create refracted rays after hitting a transmissive surface. Because of the importance regarding lens tracing, the calculation will be presented in detail. Now Snell's law is reintroduced from section 2.1.3.

$$\eta \sin\theta = \eta' \sin\theta' \tag{4.32}$$

The sine function is not gladly seen in computer graphics, but the cosine, because it can be replaced by a simple dot product of two vectors. By using the trigonometric identity it follows that

$$\cos^2\theta' = 1 - \frac{\eta^2(1-\cos^2\theta)}{\eta'^2} \tag{4.33}$$

To receive a vector from that formula, we need to introduce the refraction plane $\mathrm{P}(\vec{n}, \vec{b})$ describing a two dimensional orthonormal basis. Now we can rotate everything to that basis. The refraction direction $\vec{t}$ and the incoming direction $\vec{d}$ become

$$\vec{t} = \sin\theta' \cdot \vec{b} - \cos\theta' \cdot \vec{n}\vec{d} = \sin\theta \cdot \vec{b} - \cos\theta \cdot \vec{n} \tag{4.34}$$

which results in

$$\vec{b} = \frac{\vec{d} + \vec{n}\cos\theta}{\sin\theta} \tag{4.35}$$

After combination with (4.33), the refraction direction $\vec{t}$ can be evaluated solely with known variables [Shio5]:

$$\vec{t} = \frac{\eta(\vec{d} - \vec{n}(\vec{d}\vec{n}))}{\eta'} - \vec{n}\sqrt{1 - \frac{\eta^2(1-(\vec{d}\vec{n}^2))}{\eta'^2}} \tag{4.36}$$

The resulting equation is independent of refraction indices and ray direction. Only the normal must be orientated towards the incoming direction. The reflection case can be easily appointed with the root term, which becomes smaller than zero for the total internal reflection. The amount of refracted or reflected light is determined by the previously defined Fresnel Equations in section 2.1.3, which can be directly adopted with the extension that we do not attend to polarization. Consequently we can take the average of the terms for $T_\perp$ and $T_\parallel$.

In the moment of refraction, there is no absorption, but during passing the medium. The amount is described by Beers Law in section 2.1.

Fresnel interaction has to be sampled to receive inner and outer reflections on dielectrics at the same time with refraction. The inversion method can help out once more, by having two cases again. Because transmission and reflection coefficient T and R sum to one, it makes sense to immediately use the Fresnel transmission formulae result as refraction probability. According to (4.17) we have to weight the path contribution appropriately again. When we look to the next step we also have to lower the path weight by

multiplying with the refraction coefficient. As a consequence both operations cancel out. The only thing we have to add is the absorption during the medium transit. A path starts at location x and hits the dielectric location y. There we sample for the event $e_{Fresnel}$ whether refracted or reflected. Thus the resulting path weight w is

$$\begin{aligned} w(y) &= w(x)\frac{1}{P(e_{Fresnel})}C_{e_{Fresnel}}\exp(-\alpha|x-y|) \\ &= w(x)\frac{C_{e_{Fresnel}}}{C_{e_{Fresnel}}}\exp(-\alpha|x-y|) \\ &= w(x)\exp(-\alpha|x-y|) \end{aligned}$$

where $C_{e_{Fresnel}}$ is the valid Fresnel coefficient for the sampled event.

4.2.6 Light Tracing

In usual global illumination literature, at this place, path tracing would be described first, because it is the more common algorithm and easy to integrate in an existing ray tracing system. But the propagation of light and its trace through the scene is the more intuitive algorithm as it resembles the natural direction of propagation. The construction of light paths connecting the light source with the photo sensor is now done by sampling, as soon as more than one direction is valid. Combining all recently explained tools we are no able to solve (4.26) by (4.17) [Vea97]:

1. Sample an emission point x at a light source surface.
2. Sample a wavelength λ for the path to start and calculate the according radiance L_e.
3. Sample a correct direction ω_{o_x} for the emission.
4. Evaluate a hit-point $y = h(x, \omega_{o_x})$ and test whether we hit the camera gate/sensor, if yes, then terminate.
5. Do a Russian roulette step to probably terminate the light path at y by absorption.
6. Sample a new reflection direction ω_{o_y} according to the given BRDF.
7. continue with 4.

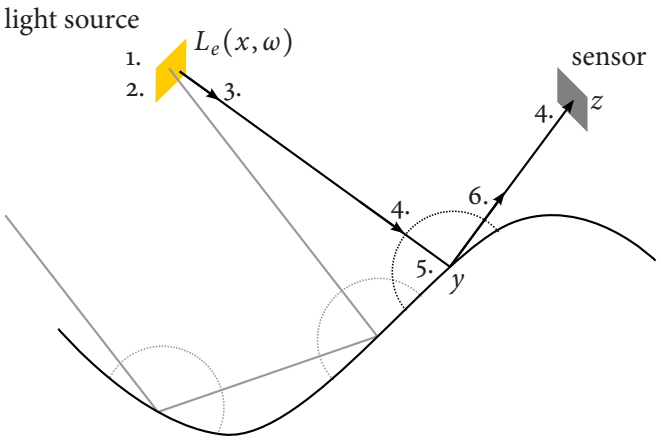


Figure 4.5: Example paths for global illumination using light tracing with sampled directions.

When starting radiance on a light source surface, we have to keep in mind, that the emission characteristic has to be introduced either by importance sampling or indirectly by uniform direction sampling and reducing the emitted radiance according to the given definition. It is the same with step 6. where the new reflection direction has to be sampled. Indeed all analytic material models which are to be used in physically based rendering already deliver the needed formulae for this step. The 5th step is needed for the reason that a light path can be arbitrarily long. If we think of two parallel standing mirrors where a light ray enters almost perpendicular to one side, the path length can become infinite due to numerical imprecision, which causes an endless path somehow. Russian roulette is a mechanism to reduce the probability of those paths, without setting a deterministic path depth, which would be wrong certainly, because we stop the estimator without sampling this event. For this method we use the inversion method, with just two cases, absorption or interaction. The probability can be chosen freely as long as they sum up to one. Obviously extreme numbers have a negative impact on the variance of the estimator. An improvemnt would be, to use information about the actual path weight value. The lower it is, the higher the probability to terminate the path.

4.2.7 Path Tracing

Backward light tracing or path tracing is the method of path construction starting at the image sensor [Vea97]. So far it was radiance that we transported along light paths, now we define an abstract unit of *importance*. We construct our rays exactly the same as in 4.2.6 now starting from the sensor and trying to hit a light source. We sample a location on a pixel, traverse the camera gate, enter the scene and multiply all path contribution changes we receive. In the end we hopefully hit a bright light source and can accumulate the emitted radiance multiplied with the path weight. This straight forward approach is based on the reciprocity theorem by Helmholtz which states, that a light source and point of view are interchangeable [Vea97], as long as all materials in the scene behave just the same way.

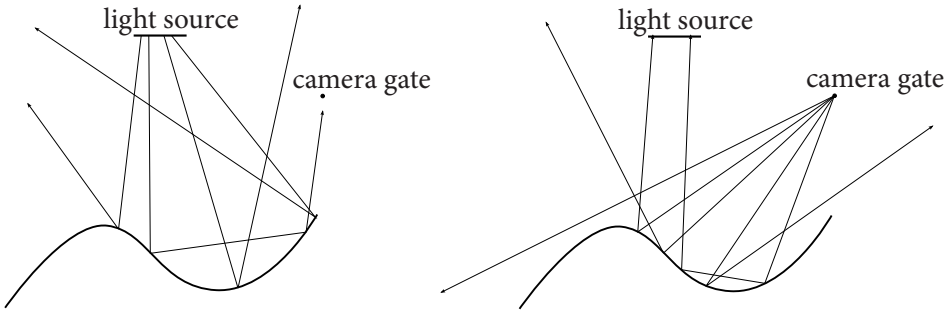
4.2.8 Direct Light Estimation

At that point we are able to construct all paths with given materials and scene setting. Unfortunately path and especially light tracing have some severe disadvantages in how probable certain paths are. This fact makes the convergence rates disastrous in certain cases. Light tracing in general has the serious flaw that sensor or camera gate are relatively small, compared to the scene. Path tracing has good convergence rates for large light sources, and is moreover ideal for scenes illuminated by an environment map. In the case of a specular event before reaching the camera, the situation gets even worse. The reflection direction then totally depends on the incoming direction and is thus additionally even harder to be sampled. It could also be said that with path tracing we sample the scene exactly in the region we look at and illuminate by secondary rays into the scene. Light tracing illuminates the scene and hopes to hit the camera.

An important step for variance reduction is based on the property of our scene description, that we know exactly where the camera gate or respectively the light sources are located. Thus, we can advance to a new path generation strategy, not by sampling directions, but by sampling surfaces again, and connecting both points. In theory, we have to reformulate the rendering equation from an integral over the hemisphere to an integral over surfaces of interest [Vea97, Shi05].

$$L(x,\omega_o) = L_e(x,\omega_o) + \int_A \rho((x' \to x), x, \omega_o) L(x', (x' \to x)) G_{xx'} V_{xx'} dx' \tag{4.37}$$

The new idea can be understood as a form of importance sampling, because we now only look in the direction we are interested in. This method is usually done for sampled points x' on the light source,



(a) Light tracing (b) Path tracing

Figure 4.6: Sampling the scene inside the field of view starting at the sensor creates much more relevant rays.

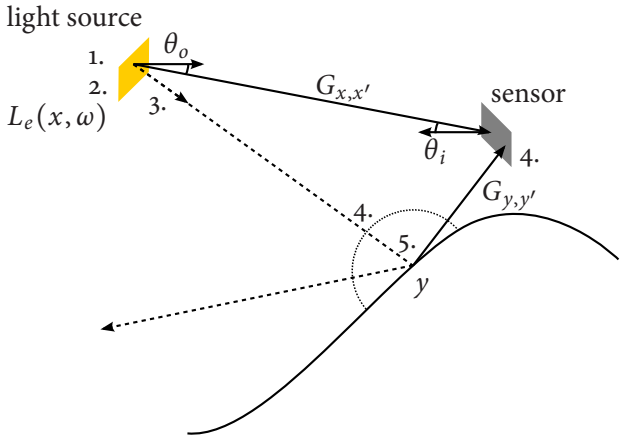


Figure 4.7: Direct contribution estimation for ligh tracing by explicitly sampling the surface of interest, the sensor in this case, concerning. In fact, reciprocity of Fermat propagation can be exploited in a way that we are allowed to swap light source and sensor, and directly receive path tracing in this graph.

respectively the camera gate, to efficiently integrate the direct contribution.

$$V_{xx'} = h(x, (x \to x')) \;==\; x' \,?\, 1 : 0 \tag{4.38}$$

is introduced as the visibility function. $G_{xx'}$ is the outcome of the integration border substitution by the change of variables theorem [BSMM05, Vea97]:

$$G_{xx'} = \frac{\cos\theta_i \cos\theta_o}{\|x - x'\|^2}. \tag{4.39}$$

Although $\cos\theta_i$ already belonged to the original formulation over the hemisphere, it is usually included here to have one term for weighting the transported radiance. The product of $V_{xx'}G_{xx'}$ can hence be interpreted as the compensation of the path weight by the factor of how probable this path is when we would have sampled it by a uniform hemisphere sample.

Since only diffuse surfaces reflect in all directions equally, this approach works best for those. For incidence depending specular or transmissive singular surfaces the method fails, because now we need to sample the destination for a Dirac pulse direction, which has zero probability.

To sum it up, the new algorithm with direct contribution estimation, this time presented for path tracing, looks like:

1. Sample a point x on a wished pixel at the sensor and a direction through the camera gate.
2. Sample a wavelength λ for the path to start and compute the starting importance.
3. Pass the camera gate and evaluate the hit-point $y = h(x, \omega_{o_x})$
4. If we hit a light source, add radiance to pixel.
5. Sample a point y' on a light source, evaluate visibility, and add radiance to pixel.
6. Do a Russian roulette step to potentially terminate the light path at y by absorption.
7. Sample a new reflection direction ω_{o_y} according to the given BRDF
8. continue with 5.

Step 4 rightly does not seem to fit in here, but is necessary to make light sources visible. Further reading is given in Section 5.3.2. What was left out until now, is the central question, how the light gets converged by the yet abstract camera gate in step 3. The upcoming sections give detailed information on that.

5 The Full Camera Model for Photo-realistic Image Generation

In the last chapter, we spoke about light transport and measuring the intensity falling onto the light sensitive layer. The image forming influence of the camera gate is yet left abstract until now. The simulation can be indeed made arbitrary complex, starting with an ideal pinhole camera, coming to the Gaussian thin lens approximation, up to the full path simulation through a realistic lens mount. The first two are now briefly discussed for comparison. There is also a fourth possibility between the thin lens approximation and the full simulation, more precisely, the thick lens approximation which is also able to approximate whole lens systems. After that, we discuss the main topic, the efficient realisation of a realistic lens system.

The given implementations assume the existance of defined operators for the class Vector3. Operators like $+ or *$ are written without seperation from scalar values. The multiplication between two vectors always implies the dot product.

5.1 Camera Models for Rendering

5.1.1 Pinhole Camera Revisited

The ideal pinhole approximation is the model of choice for prototype implementation or non-photo-realistic rendering. The principle behind is apparent. A single point in space connects the image and the object side of the camera. The ray generation is directly based on Section 4.2.8, where the center of projection is chosen with probability one. This model verifies also the indispensability of direct contribution estimation, because the probability to hit the pinhole with a sampled path direction is simply zero.

The main advantages are the simplicity and fast evaluation of the model, the result is a sharp, distortion free image with an infinite depth of field, without any physically based motivation.

5.1.2 The Thin Lens Approximation

As soon as one judges the latter model and its result as too artificial and limited, the thin lens model is the next choice, and already introduces first photographic properties. The entrance of the camera gate now has an area leading the possibility of introducing terms like irradiance. Moreover we are now able to sample a point on the entrance pupil to create the path connection between object space and image space. The name thin lens model originates from the fact that all image forming computations can be done in one plane as described in Figure 5.1.

The model operates at a given focal length and a specified distance b to the sensor. The image point O focused at with the given attributes is gained by [PPBS05]

$$\frac{1}{f} = \frac{1}{b} + \frac{1}{g}. \tag{5.1}$$

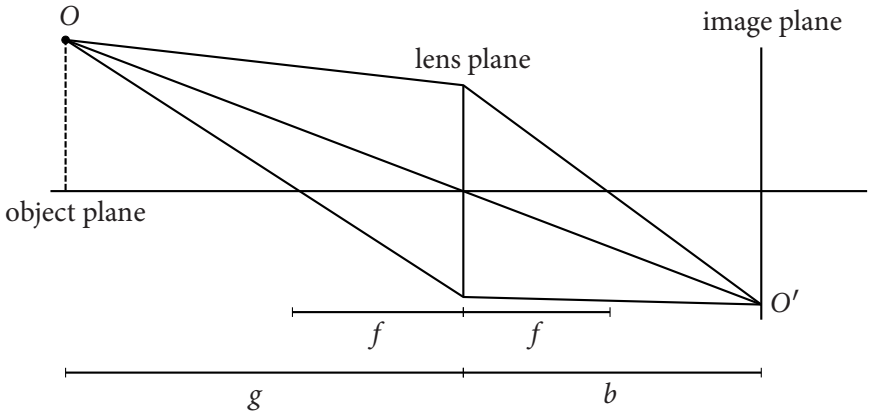


Figure 5.1: Gaussian Thin Lens Model

The first effect which catches the eye, is the finite depth of field, that reacts like realistic lenses do at first sight. The focus depth becomes shorter with higher lens radius, because the ray fan opening angle simply becomes larger.

In a nutshell, with a thin lens model, we are already able to simulate depth of field and correct exposure control by varying the f-number without bothering indices of refraction or Snell's Law. Direct contribution estimation still applies by sampling a point on the lens area and is still far more efficient than sampling directions at diffuse surfaces to hit the lens primitive.

5.1.3 The Thick Lens Approximation

This method is an extension of the thin lens model with regard to the lens thickness, described by the distance of the nodal points N_1, N_2, explained in figure 5.2. A very positive property of this approximation is the possibility to translate a whole lens system into a transformation matrix [PPBS05]. As the image formation stays linear under consideration of (3.4), this approach is valid as long as the lens geometry does not change. The result is, that a whole lens system can be approximated by one single transformation to a ray passing the camera gate. This technique is a main tool in optical design analysis [Smi07].

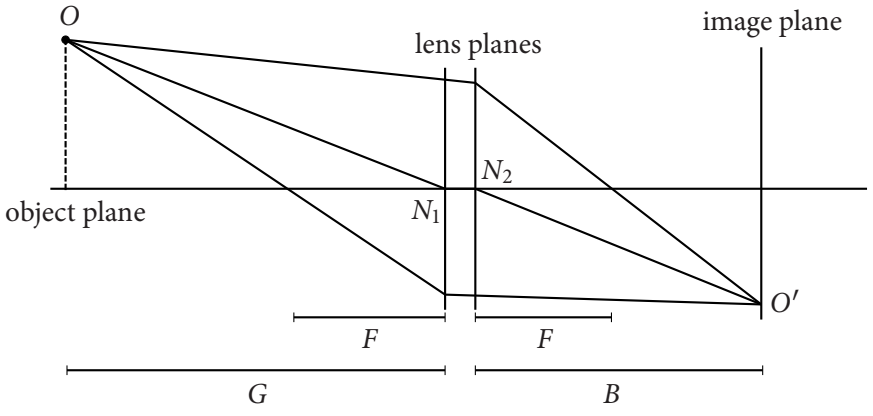


Figure 5.2: The Gaussian thick lens model introduces a thickness to the latter approach.

Although ray generation through a lens system can be hereby done efficiently, the importance in computer graphics is marginal, because of the minimal difference to the much simpler thin lens approximation. The method was not proposed for realistic rendering until 1995 [KMH95]. This paper by Craig Kolb et al. is

the basic idea of how realistic lens tracing, as it is called from now on, can be introduced into a rendering system.

5.2 Physically Based Monte Carlo Lens Tracing

Real photographic lenses are highly complex compounds which are designed under the consideration of a wealth of variables. The theoretical part on optics gave a fair overview, which will suffice for this part.

For completeness, here the example for a lens design in Table 5.1 from Section 3.2.5.

rad	thick	matl	ior	V-no	sap
42.970	9.8	LAK9	1.691	54.7	19.2
-115.33	2.1	LLF7	1.549	45.4	19.2
306.840	4.16	air			19.2
	4.0	air			15.0
-59.060	1.870	SF7	1.640	34.6	17.3
40.930	10.640	air			17.3
183.920	7.050	LAK9	1.691	54.7	16.5
-48.910	79.831	air			16.5

Table 5.1: Tessar Design by Brendel (USP 2854889)(f/2.8, 100mm EFL)

5.2.1 Lens Representation

While recalling the definition, the idea to model a real lens system with a constructive solid geometry description seems applicable, similarly proposed by [KMH95]. A concrete realisation is the combination of two spherical domes, as described in Section 4.1.1 which lead to a spherical lens element with the desired surface properties, directly adoptable from a lens design table. A further intersection with a cylinder cuts down the borders to a specific diameter. This representation has several advantages over an alternative triangle-based surface description. First, only one intersection test is needed for one surface. A triangulated dome consists of several hundred triangles, which have to be organized in an acceleration structure to keep the number of tests sub-linear.

The evaluation of the surface normal at location p on a sphere $S(c,r)$ with radius r and center coordinate c is analytically possible by

$$\vec{n} = \|p - c\| = \frac{p - c}{r} \tag{5.2}$$

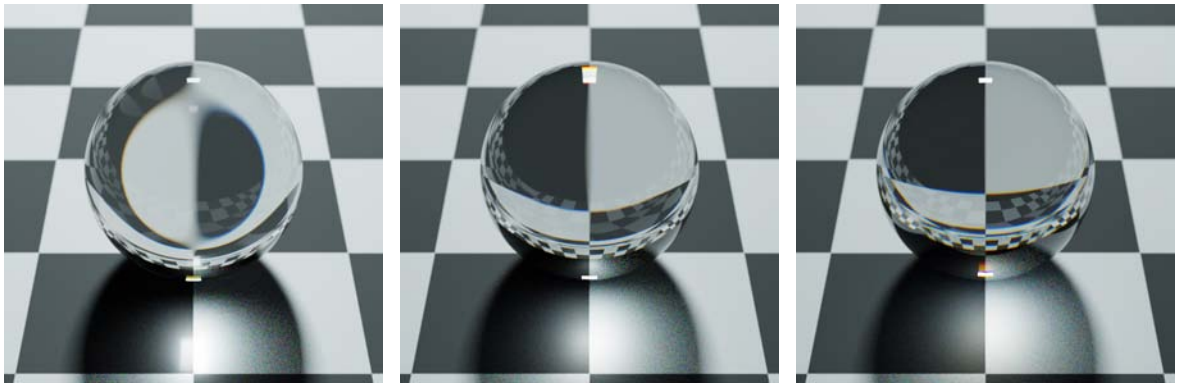
at every point. The triangle approach in general has the disadvantage, that one can only specify plane surface parts. Interpolation techniques allow to overcome this flaw to certain degree, but still cannot suffice the needs for correct refraction computation in general. This topic and some extensions to overcome this problem is also covered in detail by [Vea97]. The straight forward way for increasing the number of triangles used for approximation would indeed be correct, but also make the intersection test even slower. All possible methods are in the end too complicated to approximate such an analytially well defined surface.

Furthermore, the implementation also demands primitives for plane disc surfaces, e.g. filters, and polygons for different aperture shapes. A disc can be described as a plane with and according radius limiting the valid area around the center.

Besides the usual scene environment, a lens environment was set up in CHROMA as an isolated region, where all surfaces are stored in order. Ray generation is sourced out to to the lens tracing module, a reduced ray tracing kernel that is able to handle the specific needs of path generation through the imaging system.

5.2.2 Glass Data

The lens design description also includes the specific type of glass needed, to achieve the intended optical performance. The Schott Glass catalog exists in electronic available at `www.Schott.com`, from where the Sellmeier coefficients can be extracted to have the correct material description.



(a) L-CAF2:Low IOR, high Abbe-Value. (b) N-LAK12: Mid IOR, mid Abbe-Value. (c) N-SF66: high IOR, low Abbe-Value.

Figure 5.3: Different types of glass rendered with CHROMA. Specifications taken from [Sch09].

Under consideration of the former explained concepts of spectral ray tracing and Snell's law, the general layout for a *lens tracer* is relatively straight forward. All further simulation aspects are now summarized in the following section.

5.2.3 Ray Generation

With the following approach, tracing rays through a real lens system can be done with exactly two tools, an intersection test for every primitive used, and a general function for the refraction evaluation. The next intersection event can be determined solely by exploiting the knowledge of surface order and the fact that refraction is a deterministic process. That makes an acceleration structure obsolete. The implementation for tracing a ray through a lens system is given in Algorithm 2.

The proposed top-level core of the CHROMA lens tracer hides some special properties implemented. The method is called, once it is sure that a ray enters the lens system. By inserting a trigger object to the scene for the light tracing case, we connect the world and the lens and can thus handle both separately. This trigger primitive can be a disc as large as the front most lens surface for example which can be also used for direct light estimation. For the moment, the path tracing approach samples the back most surface for a deterministic connect from the sensor to the lens.

The separation allows for having two different coordinate systems, one for the scene and one for the camera-internal computations. To minimize numerical imprecision, a change to dimensions of millimeters is recommended, if realistically dimensioned scenes are to be rendered. A transformation to a wished camera coordinate system can also simplify vector calculations, as for example the ray-plane intersection. Figure 5.4 gives a visual insight how CHROMA handles ray processing.

Algorithm 2: Lens Tracing Kernel

```
   Data: ray from scene or sensor side with given path weight
   Result: leaving ray, actual path-weight and ray sign according to optical axis
 1 int testIndex;
 2 if ray.direction[axis]>0 then
 3 |   testIndex = firstElement;
 4 |   raySign = 1;
 5 else
 6 |   testIndex = lastElement;
 7 |   raySign = -1;
 8 end
 9 float attenuation;  while firstElement<=testIndex && testIndex<=lastIndex && pathweight>0 do
10 |   hitpoint = intersect(ray,surface[testIndex]);
11 |   switch hitpoint.hit do
12 |   |   case lens
13 |   |   |   attenuation = refract(ray,hitpoint);
14 |   |   end
15 |   |   case aperture
16 |   |   |   attenuation = diffract(ray,hitpoint);
17 |   |   end
18 |   |   otherwise
19 |   |   |   pathweight = 0.0;
20 |   |   |   return 0;
21 |   |   end
22 |   end
23 |   testIndex+=raySign;
24 |   pathweight *= attenuation;
25 end
26 return raySign;
```

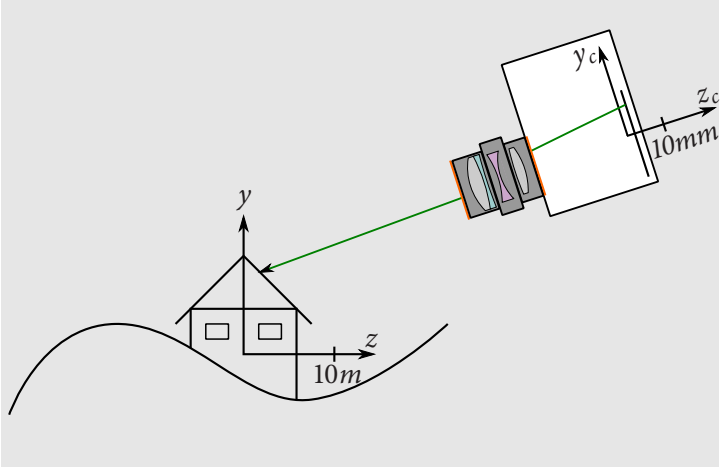


Figure 5.4: In the white camera area, a ray is formed by connecting lens system and sensor either deterministically or by direction sampling. This can be applied to all models. The ray generator seen as black box by the rendering system deliver the appropriate leaving direction for further scene-based path tracing. The orange line symbols the trigger gate for path hits coming from the scene.

The effective path and its weight is determined by sampling techniques for path construction from Section 4.2.5 and Section 4.2.8, including Fresnel based refraction and reflection from equations (2.5), Snell's Law (2.2), and Beers Law (2.6).

Let us briefly complete the path construction beginning at the sensor. On a given pixel p_{ij} we choose a uniformly distributed random location x to start the ray. It is common to take one sample for one every pixel per rendering pass. Thus we took samples in the quantity of resolution, all uniformly distributed on the sensor, resulting in the sample probability $P(x) = \frac{1}{A_{sensor}}$. The direction is gained by uniformly sampling the first surface of interest with area $A_{surface}$ yielding y. Hemisphere sampling is horribly inefficient, as the lens only covers a small solid angle seen from the sensor. From previous explanations, and Section 4.2.8 as well, we can conclude that the path weight w for the formulated ray from x to y is:

$$\begin{aligned} w(y) &= \frac{1}{P(x)} \frac{1}{P(y)} G_{xy} \\ w(y) &= \frac{1}{A_{sensor}} \frac{1}{A_{surface}} \frac{\cos\theta_x \cos\theta_y}{\|x-y\|^2} \end{aligned} \tag{5.3}$$

The angles in the geometric term are meant to apply to the equivalent surface normal. The eligible question which surface to sample in detail is delayed until the next section. For the moment let us be comfortable with just sampling the back lens. The implementation is given in Algorithm 3.

This approach can now be directly extended by simulation of material coatings. It is assumed that the layer itself has no further influence than changing the Fresnel coefficient according to the path wavelength. This is not completely correct, as interference effects play a certain role here, but this topic has been neglected before. After recalling Figure 3.13 it is clear that we could directly apply the relative change of reflection degree to the actual received reflection value for a specific wavelength. This has no influence on the ray path, only on the path weight, which finally determines the amount of transported radiance. Consequently, the thickness of the coating must be zero and has instantaneous effect at the dielectric location in this

approximation.

Algorithm 3: Lens Surface Refraction

```
   Data: Ray and hit-point on dielectric surface
   Result: new ray direction form hit-point and attenuation
 1 Material* matTo, matFrom;
 2 hitpoint.getMaterials(matTo,matFrom);
 3 float etaFrom = matFrom->getIOR(ray.lambda);
 4 float etaTo = matTo->getIOR(ray.lambda);
 5 float cosTheta = ray.dir*hitpoint.normal;
 6 float sqrtTerm = 1.of - ( etaFrom * etaFrom *(1.0-cosTheta*cosTheta))/(etaTo*etaTo);
 7 float weight = 1.of;
 8 Vector3 newDir;
 9 if sqrtTerm>0.of then
10 |   newDir=((etaFrom*(ray.dir-hit.normal*(cosTheta)))/etaTo-
   |   (hit.normal*sqrt(sqrtTerm))).normalize();
11 |   double refractance=getFresnelTransmission(cosTheta,newDir*hit.normal,etaFrom,etaTo);
12 |   matTo->coating(ray.lambda,refractance);
13 |   double brdfCase = sampler.getRandom();
14 |   if brdfCase > refractance then
15 |   |   newDir = reflect(ray,hitpoint);
16 |   |   weight = (1.of-refractance)/refractance;
17 |   end
18 else
19 |   newDir = reflect(ray,hitpoint);
20 end
21 ray = Ray(newDir,hitpoint);
22 return weight*matFrom->getAbsorption(ray.length,ray.lambda);
```

Linear interpolation on a data set with support nodes every 50nm already deliver adequate results, as seen in Figure 5.5.

As long as we pass all lens surfaces, the loop carries the path to one of both exits by the condition to have a valid index for testing. If the ray gets blocked during transit, the switch returns a miss. Otherwise, if the loop terminates correctly, we return the correct ray sign and appropriate path weight. This implementation can be easily extended to a scenario with barrel flare. By replacing the default case, which is entered when missing the next surface, with a reflection event on the lens barrel cylinder, we again receive a new direction and attenuation according to the defined BRDF.

5.2.4 Aperture Stop Simulation with Diffraction Account

Aperture stop simulation means that we insert an opaque type of surface to the lens tracing system. The implementation handles it by defining a special primitive for apertures, which has some additional possibilities besides scene primitives. We are allowed to resize the radius or change the shape by overwriting the base primitive it points to. The model used in CHROMA allows to use a disc, or a circle-approximating polygon. Even a quadratic binary texture, for profile definition could be used. The inclusion itself into the lens tracing kernel can be done by inheriting from the simple surface primitive, thus we have no *if* and no further changes to make in Algorithm 2, except to check for the hit code, anyway done by the switch

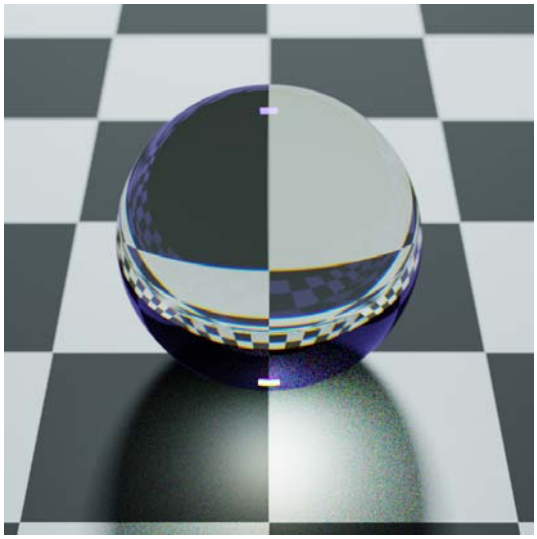


Figure 5.5: Glass type N-SF66, already seen in Figure 5.3, now with an additional double layer V-Coating with values taken from Figure 3.13 which reduces reflection around 550nm beyond 1% but has almost no effect to the lower and upper end of the visible spectrum.

case block. This avoids an unnecessary call of the refraction routine, which would be valid indeed, if the material of the aperture object is set to *air* for example.

It is important to note that, even without an aperture, the effectively sampled area by paths through the system is the plane where the cone of illumination has smallest diameter limited by the lens diameters itself. In a real lens, the aperture explicitly defines this plane as already explained in Section 3.2.3. Rays that become blocked at the diaphragm are simply absorbed.

Diffraction

The topic of diffraction is yet a rather less covered topic in computer graphics, as only very few approaches were published that do not introduce these effects by a post process. In our setting, we are interested in diffraction effects in one place, at the aperture gate.

Nevertheless, the following proposals are based on a paper by Lilian Aveneau and Michel Mériaux [AM99], who published an algorithm and data structure to solve diffraction account in polygonal scenes. In the following, I present their used diffraction coefficient, and their method for finding the diffraction path PQR for a given source location P and destination R. Keep in mind that previous work from Section 2.1, and also the upcoming, only cover diffraction at a straight edge. Thus we are limited to polygonal apertures. There exist further geometrical diffraction models for curved surfaces and vertices not considered so far. For appropriate theory, I once more recommend [Kel62]. Transfers of these cases to computer graphics were not known until the date of completion of this thesis.

The diffracted radiance L_d for a diffracted ray of unpolarized incident light L_i can be expressed by:

$$L_d = L_i \lambda \left(\frac{\cos\frac{\pi}{4}\sin\frac{\pi}{n}}{n\pi\sin\beta\left(\cos\frac{\pi}{n} - \cos\frac{\delta-\gamma}{n}\right)} \right)^2 \tag{5.4}$$

Here γ and δ are the angles between the wedge plane normal and the projections of the rays' directions into the plane through Q that includes the normal [AM99]. For a better understanding, Figure 5.6 can

help, and Figure 2.5 is given there once more in an extended version. The resulting radiance can be added to an existing result without harming energy conservation laws.

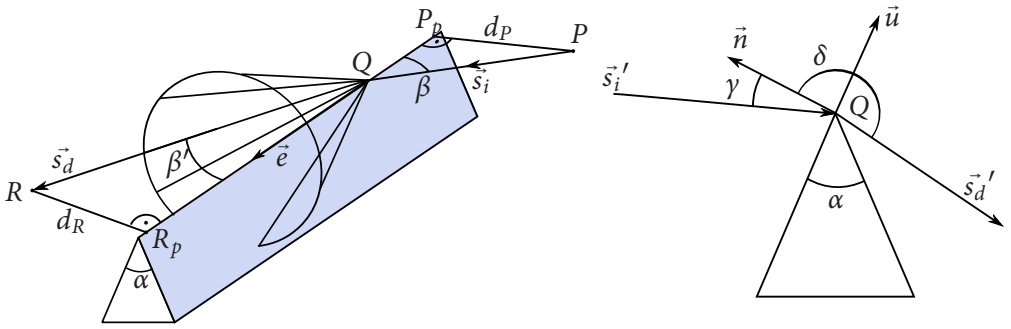


Figure 5.6: The left drawing is an extended repetition from Section 2.2.2. P_p and R_p are projections of source and destination onto the edge. On the right, a two dimensional projection into the plane perpendicular to the edge vector $\vec{e}$ through Q is shown. Two new angles, γ and δ are introduced, which lie between the normal of the hit wedge plane and the projected vectors $\vec{s}_i'$ and $\vec{s}_d'$. Direction $\vec{u}$ completes the *edge coordinate system* together with $\vec{e}$ and $\vec{n}$

The diffraction path construction, given in [AM99], assumes that the points P and R are already known. Now let t_{P_p}, t_Q and t_{R_p} be the parametric ray coordinates of the points P_p, Q and R_p. The edge is determined by its vertices A and B. Q can be evaluated by the following geometric approach, once at least P_p or R_p lies between A and B. Based on $\beta = \beta'$,

$$\begin{aligned} \tan\beta &= \tan\beta' \\ \frac{d_R}{(t_{R_p} - t_Q)} &= \frac{d_P}{(t_Q - t_{P_p})} \\ t_Q &= \frac{t_{P_p} d_R + t_{R_p} d_P}{d_P + d_R}, \end{aligned} \tag{5.5}$$

with $|P - P_p| = d_P$ and $|R - R_p| = d_R$, as given in Figure 5.6. Q is valid if $t_Q \in [0, |B - A|]$.

When recalling Algorithm 2 we realise that we have only one point given, and this is P, representing the starting point on the last lens surface before the aperture. The reason, why this approach works in scene environments is the possibility to sample points on surfaces, where we would like to compute the diffracted radiance that reached this location from the source. A direct light estimation setting could hereby be extended with a second chance for rays that become blocked by obstacles between surface sample and light source. [AM99] introduces a data structure to search for a diffraction point in the scene inside a defined ray tube, to connect these points by a diffraction step. Unfortunately this operation is only valid for non-singular surfaces, because the search algorithm determines a new direction towards a diffraction point, which has automatically zero contribution in our setting with singular dielectric surfaces. This approach can only be applied for the case, that the source point lies on a non-singular surface, like for example the sensor. Assumed that the aperture is the last system surface, in a path tracing setting we could then connect a given pixel sample with a location sample on the first glass surface by the given method if no direct path exists. Unfortunately a setting with a back aperture is rather unusual, and thus the method stays here for later comparison with further ideas.

Now let us restart and develop a new approach under the consideration that we need to build up a path of the form TDT where D is a non-singular diffraction location in the aperture plane. A straight forward idea is to use the same tools like for BRDF evaluation, sampling a new direction and weighing it with the given probability and appropriate BRDF modifications. The comparison already suggests a nice realisation.

In CHROMA, I implemented a diffraction material that can be applied to a primitive. The method of operation is directly applicable from known materials. Given a hit-point and incoming ray, we can return a new direction and according path weight. In the setting of diffraction, both can again be obtained by known techniques.

The cone of diffraction represents all valid directions for a diffraction event at Q. Uniformly sampling a point S on the base circle perimeter of the Keller cone yields a direction $\vec{s_d} = (S - Q)$ with according probability $\frac{1}{2\pi r_{circle}}$. Yielding the outgoing direction, we instantly have all needed angles to solve (5.4).

This solution also has two serious flaws. To keep the setting of [AM99], the hit-point must lie on the aperture edge. Hence the surface is now singular with respect to the possibility to hit the structure. The attempt to sample an edge or even a point with a random direction in space is equal to sampling a Dirac pulse again, which has zero probability. A possibility to yield an approximate solution is to execute a diffraction event, once we *almost* hit the edge. In fact, diffraction comes up, as soon as wavelength and distance to the obstacle approximate. Thus I made experiments considering the evaluation of the barycentric coordinates which can be gained during the appropriate intersection test. As soon as α, β or γ are almost one, the hit-point is near the according vertex v_0, v_1 or v_2. For edge proximity, the sum of the associated barycentric coordinates per vertex reaches one. Also compare the suggestions with Figure 5.7. Because pure checking of β and γ

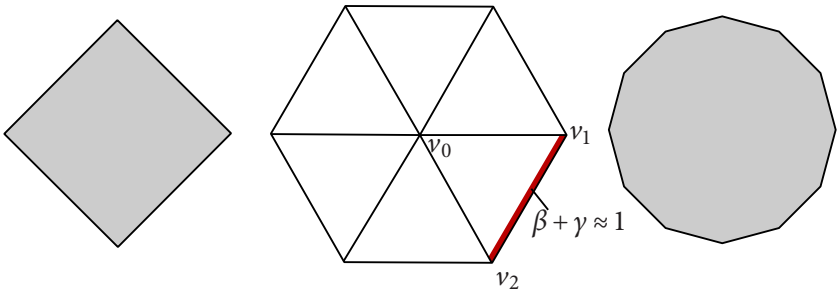


Figure 5.7: Overview of the tested aperture shapes. The proposed barycentric coordinate test is outlined for one example. Here,the region of interest is marked red. Hit-points in there, are used to estimate diffraction contribution. The region will further not be used to let primary image information pass. Thus, the aperture radius is reduced a bit.

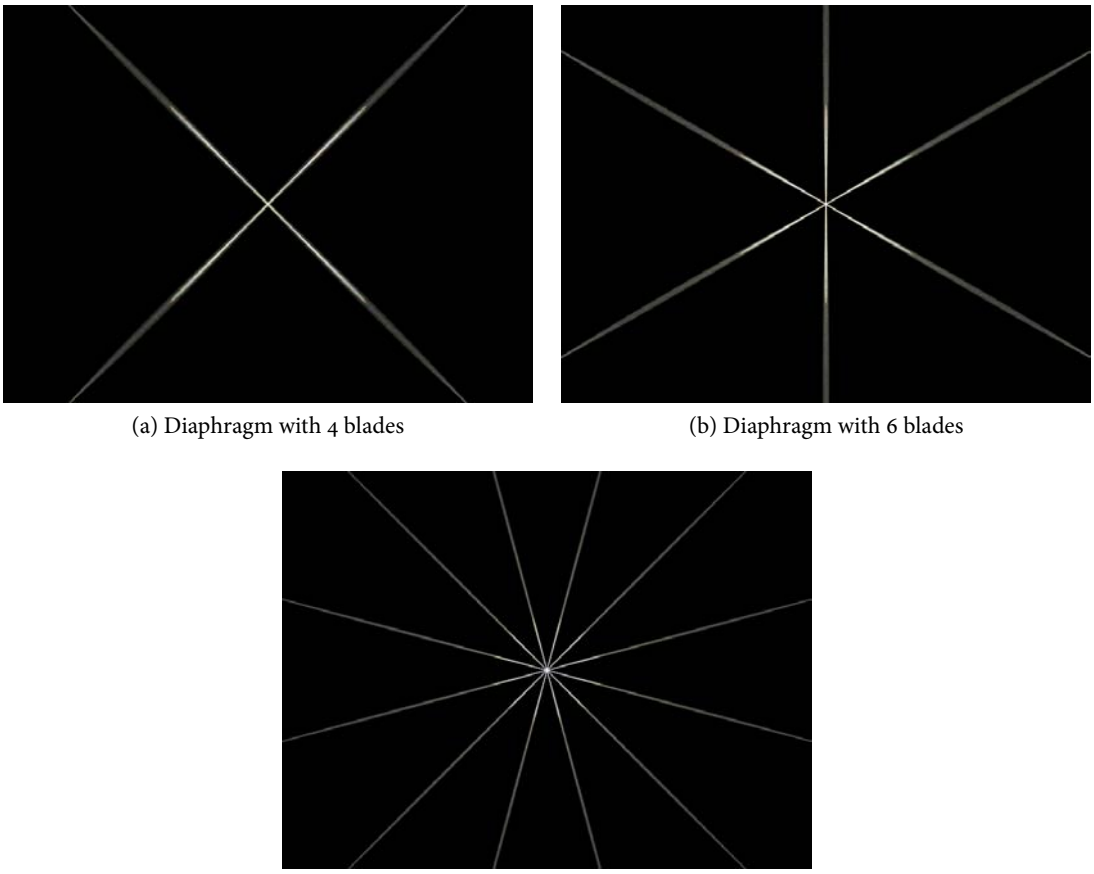
does not account for aperture dimension, we need to introduce the absolute value to the comparison. In my implementation, I chose a diffraction threshold of 0.99 for unit aperture radius, which will be increased for larger apertures, and decreased for smaller ones respectively, by

$$th_{diff} = 1 - \frac{0.001}{r_{AP}}. \tag{5.6}$$

The resulting idea is to compute diffraction directions for all those rays, which lie in the diffraction zone of the aperture. The method reduces the diameter of large apertures by a negligible part, but becomes problematic for very small gates like pinholes with diameters less than 0.01mm, where the inner image forming part of the aperture rapidly runs to zero. Algorithm 4 sums up all the just introduced ideas for a plane, infinitesimally thin aperture with wedge angle $\alpha = 0°$ or $n = 2$, according to (2.7). The given code does not raise claim for efficiency. I removed all performance increasing elements for the sake of comprehension.

A satisfying conclusion to this topic cannot be given. The primary intention was to set up a basic approach to simulate diffraction effects at the aperture gate, making the formation of diffraction patterns possible. The relatively novel topic of Geometrical Diffraction Theory demands a deep practice of wave optics and field theory. The proposals by Aveneau and Mérieaux are, as far as my search brought up, the only ones made, to simulate diffraction events in computer graphics in a ray tracing setting. I would like to give

some further notes on questions that arose to me during my examinations. The diffraction coefficient given by [AM99] does not account for the distance to the diffraction point, like the classic formula (2.8) does. An explanation based on the disregard of polarization and phase is not visible to me. Moreover, the unit of λ is not given for (5.4). Both mentioned things were not addressed. The following results are thus definitely not correct when considering absolute resulting illumination by diffraction. On the other side, the received pattern by sampling the directions along the Keller cone already look familiar, when compared to the photographs in Figure 5.10. Results achieved with my proposed rendering algorithm are presented in Figure 5.8.



(a) Diaphragm with 4 blades

(b) Diaphragm with 6 blades

(c) Diaphragm with 12 blades

Figure 5.8: The scene only consisted of a bright point light source. The results clearly show a propagation direction of the streaks perpendicular to the diaphragm edges.

Finally it is to mention, that Keller's original approach is not valid for all directions of diffracted rays. The regions along the red lines in Figure 2.6 marked as ISB (illumination shadow boundary) and RSB (reflection shadow boundary) result in angles so that the diffraction coefficient becomes infinite [MPM90]. This property can be eliminated by the uniform geometrical diffraction theory approach introducing a more general diffraction coefficient and some additional approaches. By sampling the diffraction direction, practically we do not need to worry, because this time it is our advantage to be unable to sample these singularities. But we come arbitrarily close, which means the general solution should be considered in future work.

Algorithm 4: Aperture Diffraction Method

```
. Data: Ray and hit-point on diffracting surface
  Result: new ray direction form hit-point and attenuation
if hitpoint.u + hitpoint.v < 1.0-DIFFREPS/aperture.radius then
    newDir = ray.direction;
    return 1;
Trianlge* hitAPTriangle = hitpoint.hitPrimitive;
Vector3 edgeV = (hitAPTriangle->p2 - hitAPTriangle->p1).normalize();
Vector3 P = ray.origin;
Vector3 Q = hitpoint.location;
Vector3 sI = (P-Q).normalize();
float cosBeta = sI*edgeV;
float beta = acos(cosBeta);
if cosBeta<0 then
    cosBeta *=-1;
else
    edgeV *=-1;
end
double u = sampler.getRandom()*2*PI;
Vector3 perimeterSample = Vector3(cos(u),sin(u),0);
transformToEdgeCoordSystem(edgeV,perimeterSample);
float coneLength = cosBeta/sin(beta);
float baseRadius = sin(beta);
Vector3 R = Q + coneLength*edgeV + baseRadius*circSample;
Vector3 newDir = (R-P).normalize();
float probability = 2*PI;
float gamma=0, delta=0;
Vector3 Pproj = P - (Q*edgeV - P*edgeV)*edgeV;
Vector3 sIproj = Pproj-Q;
gamma = acosf(sIproj.normalize()*hitAPTriangle.normal);
Vector3 Rproj = R - (Q*edgeV - R*edgeV)*edgeV;
Vector3 sDproj = Rproj-Q;
delta = acosf(sDproj.normalize()*hitAPTriangle.normal);
float cubeTerm = cos(PI/4)/(2*PI*sin(beta)*(-cos((delta-gamma)/2)));
float diffrWeight = ray.lambda * cubeTerm * cubeTerm;
ray = Ray(newDir,hitpoint.location);
return diffrWeight*probability;
```

Figure 5.9: Final example images with a diffraction layer, computed with the presented methods. The result has a significant impact on brightness perception of the light source spots.

Figure 5.10: Real glare streaks formed by aperture edges. On the left side, above the lantern, two main streaks can be identified. On the right side, there are many short streaks, arising from the border ring of the spot light. But also main streaks in all directions are recognizable. Also compare the given examples with Figure 3.17, where the shot towards the sun produces large streak bands. Left photo by courtesy of Katharina Schwarz.

5.3 Improvements

It is time to give an answer to the question *where* exactly to sample finally. We can choose random locations everywhere as long as we know the appropriate surface of the sampled region imaged into the aperture plane. It follows that the path weight for a uniform sample on the back or front lens, to construct a connecting ray, is not the lenses' area but the area of the projected aperture. This independence is based on the transformation character of lenses. Unfortunately the sampled aperture area can vary with pixel position, already known from section 3.2.7 and as seen in the following.

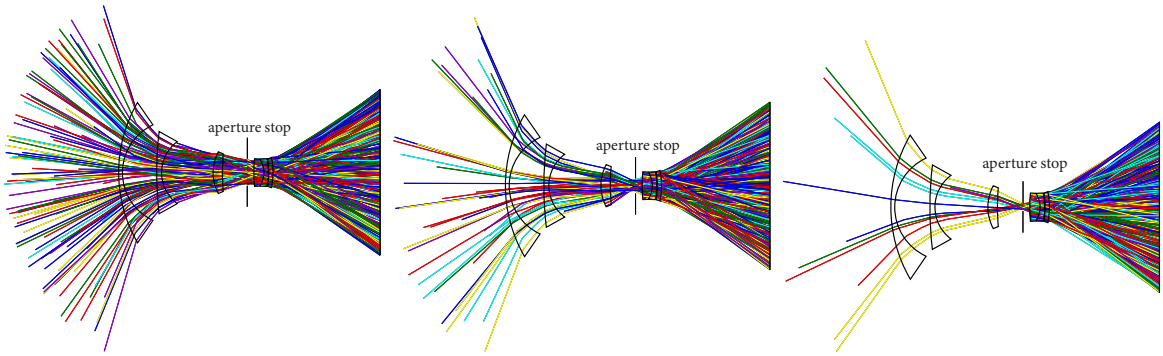
Reflected rays inside the lens can take an arbitrary path to the sensor, depending on ray height and slope, which also means that they can generally leave the back lens everywhere. I further call all contribution by paths including at least one lens reflection, *secondary image contribution*, because artifacts thus evolve from a secondary effect. The incidence of reflection can be cancelled out by not sampling reflection events inside the lens, or giving reflection zero probability. Total internal reflection due to reaching the Brewster

angle does also have zero contribution to primary image formation. The following section only applies to images generated by primary image information, solely gained by refraction without diffraction paths.

5.3.1 Pupil Sampling

Analysis

Before I explain the basic problem in words, study Figure 5.11 where the upcoming is outlined visually. By plain sampling at the back lens surface, we produce many paths that are not able to reach the other side of the lens. This becomes obvious, once we take a look at fisheye photographs, where the pupil diameter is not even able to cover the whole sensor surface. This is a special property of that type of lens, in order to reach the desired tremendous field of view. Over 45% of all pixels never receive contribution when we assume a $36x24mm$ sensor and an ideal circular image. Nevertheless in the naive approach we started rays on all pixels.



(a) at $f/4$ (b) at $f/8$ (c) at $f/16$

Figure 5.11: Lens trace plot of a fisheye lens (design by Muller) [Smi05] at 14mm focal length. Back lens surface samples for a starting direction lead to many zero-contribution rays, which the system is not able to transfer to the object side.

At this point I introduce a technique to CHROMA, I call *Pixel-Specific Pupil Sampling*. It evolves primarily from the just explained problem, and furthermore from the observation, that there indeed exists a uniquely sized pupil for every point on the image plane, caused by the varying effective aperture. In the special case of a fisheye, the regions not illuminated have a pupil with zero radius. [KMH95] proposed similar projection ideas based on Gaussian approximation, which is not able to handle all types of lenses. Furthermore, it is not made clear, how the variation of the exit pupil size over the film plane is considered exactly.

Construction

In CHROMA, I first defined a further virtual surface, comparable with the object side entrance gate, but this time on the sensor side. This plane determines the *global* exit pupil formed by the cone of illumination and the intersection of that plane location. Its position along the optical axis should be directly behind the last surface. Keep in mind, that generally, pupils constructed by optical engineering rules also have fix positions defined by previous rules in Section 3.2.3. Though the term pupil describes the same thing, our construction rule is different as we wish it at the specified location.

The first stage is to evaluate the global exit pupil, offering the maximum valid pupil for all positions. Directly sampling the constructed global exit pupil for path construction does not give any noticeable improvement

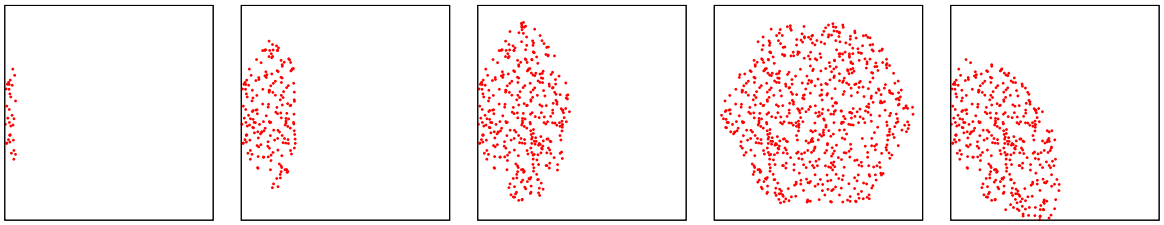


Figure 5.12: Sampling the image of the effective aperture for different pixel locations. The pixel pupils shows large distortion. The first four images show the positive sample locations in the pupil plane for pixel locations on the mid pixel row, starting at the fisheye characteristic illumination border ending at the center pixel of the sensor. The last image depicts the shape for a central pixel in the first quadrant of the sensor.

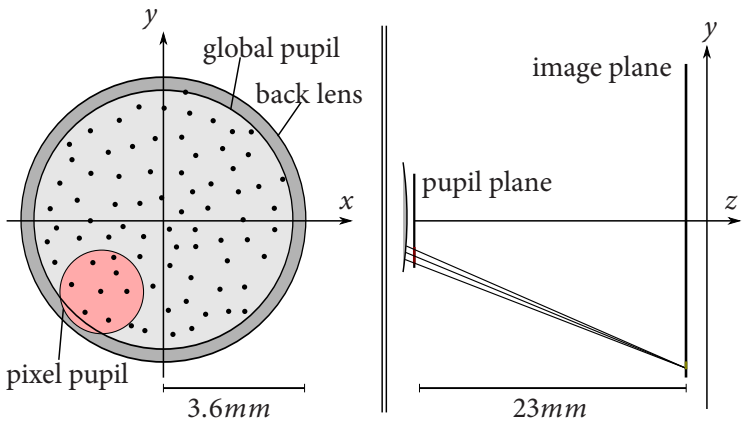


Figure 5.13: Samples of ray directions on the global pupil are stored if the constructed path reaches the object side. The average of the positive samples yields the center for a pixel pupil. The example is based on real computed pupil data for a 14mm Muller fisheye lens. The used pixel has position $-12x - 12.75$mm on the sensor, which lies in the vignetting region on the image boarder. All dimensions are proportionally correct. Because the circle shape is rather imprecise there is still a negative region off the pupil boarder which still causes a small but negligible amount of zero-contribution rays.

in general, because the lens design diameter is already chosen according to the resulting exit pupil at full aperture. But as soon as the lens is stopped up, the global pupil can change and reduce the number of blocked rays at the aperture.

Before we come to the second stage, let me reconsider the second aspect why we indeed need to sample another region than the obvious full back lens surface. An additional reason for the critical rate of blocked rays is optical vignetting, which arises in marginal regions of the sensor. As already outlined in 3.2.7 the effective aperture area decreases with ray obliquity. The crux finally is that we are able to compute this effective aperture *per pixel* in the form of a unique pupil, which can be understood as image of the aperture seen from the actual pixel. Figure 5.12 depicts an example of possible pupil change over the back lens surface.

A straight forward method for evaluating the pixel pupil is to choose samples at the global exit pupil and to trace these constructed rays onto the object side. If successful, we store this sample point. After a certain number of samples we can estimate the pixel pupil according to the successful passes. A fair definition can be given by a disc. The sample points can be averaged to obtain the center, a look on the farthest positive or

Muller Fisheye 14mm	passing rays/frame		
Method	$f/4$	$f/5.6$	$f/8$
GPS	26.0%	17.8%	10.2%
GPS2	43.2%	29.7%	16.9%
PPS	64.4%	76.2%	79.7%

Kimura Wideangle 36mm	passing rays/frame			
Method	$f/2$	$f/2.8$	$f/4$	$f/5.6$
GPS	41.8%	25.4%	13.1%	7.7%
PPS	82.2%	84.5%	85.4%	85.7%

Brendel Tessar 50mm	passing rays/frame		
Method	$f/2.8$	$f/4$	$f/5.6$
GPS	60.7%	34.2%	18.0%
PPS	81.6%	83.7%	83.7%

Kingslake Tele 250mm	passing rays/frame		
Method	$f/5.6$	$f/8$	$f/11$
GPS	22.4%	15.8%	10.7%
PPS	86.7%	86.2%	86.2%

Table 5.2: Rate tables of passing rays for various lens types, averaged over all pixels. Designs excerpted from [Smi05, Smi07]. GPS stands for global pupil sampling, PPS means pixel pupil sampling. GPS2 in the fisheye table describes the rate corrected by the factor of rays which cause a blocked path by sampling pixels outside the illumination region. The still existing number of blocked rays are caused by the crude approximation of the polygonal aperture by a too large disc. Optical vignetting and considerable distortion of the pupil cause the variation of passage rate over the different aperture settings. Widely openend lenses lose passage rate in the marginal regions, because the whole aperture is not visible, thus the disc is chosen too large. Additionally, distortion deforms the pupil shape as can be seen in Figure 5.12, and a disk again becomes less suitable to approximate the pupil.

nearest negative sample from the center can deliver an appropriate radius. An illustration is given Figure 5.13. The number of passing rays, which has a noticeable impact on convergence, is thus increased by a factor of up to 8, a short comparison follows in Table 5.2.

During pupil evaluation we indirectly evaluated another important number, which allows us to define the path weight for sampling the new pixel pupil. The aperture image seen from a pixel locations can be cropped by limiting lens diameters. Exactly the same applies to all rays, especially those we took for the pixel pupil evaluation. Consequently we can use the relation between the number of passed rays n and the total number of tested rays N per pixel as exactly the same relation between effective and real aperture area. It can be verified that

$$\frac{n}{N} \approx \frac{A_{AP,effective}}{A_{AP,real}} \tag{5.7}$$

for a sufficient number of samples per pixel according to the lens system. Evidence for the relation can be given by the experiment in Table 5.3.

Brendel Tessar 50mm	$f/2.8$	$f/4$	$f/5.6$	$f/8$	$f/11$	$f/16$	$f/22$	$f/32$	$f/45$
passing rays/frame	64.31%	35.8%	18.1%	9.0%	5.1%	3.0%	1.7%	0.8%	0.4%

Table 5.3: For this experiment, I chose the specific pixel $(i, j) = (128, 96)$ on a $512x384$ sensor. After overcoming the optical vignetting charged full aperture setting, the rate of passing rays during global pupil sampling approximately halves while stopping up. That is the same rate of change, as the real and effective aperture area are subject to with every stop up. The differing rates of change from full aperture to one stop up, also found in Table 5.2, thus must describe the optical vignetting contribution to the change of effective aperture area, as no other part aside the aperture produces blocked rays. If we assume $f/45$ to be free of optical vignetting and double the percentage for every stop down, we receive approximately 100% at full aperture, which would be correct in a setting free of optical vignetting.

The construction method offers some simple improvements and benefits that can be implemented:

1. The pupil data can be stored for reuse.
2. The pupil computation is easy to parallelise.
3. If the lens is rotational symmetric, we only need to evaluate one sensor quadrant. We could mirror this data along the axes to the other quadrants for the price of bias due to implicitly reused sampling pattern.
4. The description of the pupil by an ellipse is much more precise, because the effective aperture shape for off-axis rays looks like a skewed circle. Thus the false region in Figure 5.13 outside the global pupil would be minimized.
5. An even more precisely evaluated pupil, for example by a cons tructed polygon furthermore allows to omit the aperture intersection test for reflection-free image formation, because we calculated the valid cone of illumination in advance, and only this part is then surely sampled.
6. For a fisheye lens, pixels without contribution, precisely zero pupil radius, can be identified and masked out before ray construction.

What we implicitly revealed is the additional possibility of canceling out the effect of vignetting. We could sample the new pixel pupil and use the real aperture area as constant weight for all positions. Figure 5.14 shows a visual verification of this fact.

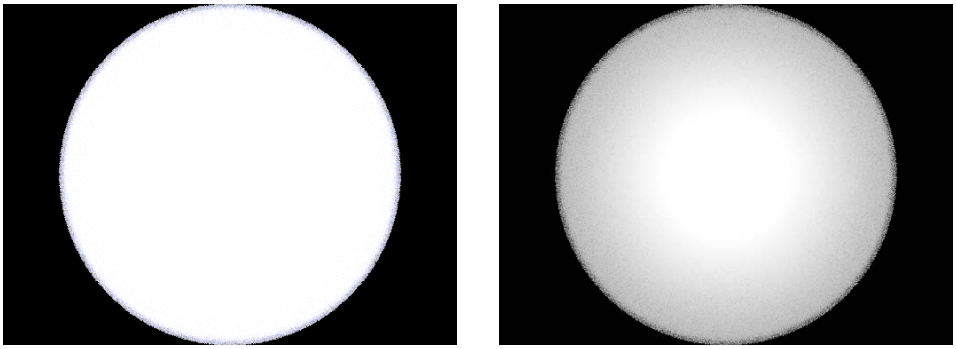
Correctness

The correctness of the pupil highly depends on the number of used samples for pupil determination. Surprisingly, image artifacts due to a too small estimated pixel pupil only result in faint or even no image artifacts, as seen in Figure 5.15. Too large pupils are no problem at all, because paths through samples in the *negative* region result in blocked rays at the aperture or the lens body.

As soon as too few sample paths are chosen, the estimated effective aperture area has a high deviation from pixel to pixel, which results in unequal brightness seen as noise typically on flat surfaces. Unfortunately this image bias does not completely vanish, as long as we use the same pupils for the whole rendering pass. For the correct solution we should recompute the aperture area every accumulation, but experiments showed, that a re-computation every 100 frames gave superb results.

The seemingly nice improvement of reusing the same set of samples for every pixel can result in strange illumination artifacts. Thus especially grids and other regular sampling strategies should be avoided.

During my experiments, I used 128 pupil samples, which means, the pupil construction time for one quadrant is in the range of seconds for 1024x768, resolution with appropriate extensions like multi threading and mirroring along the axes.



(a) Example neither optical nor natural vignetting

(b) Natural Vignetting

(c) Optical vignetting

(d) Both vignetting forms in one image.

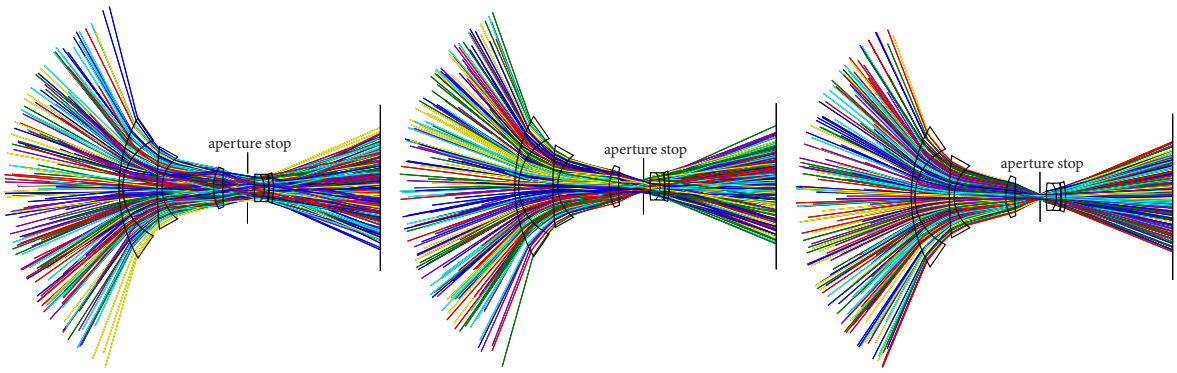
Figure 5.14: Rendering an equally illuminated grey-card makes vignetting clearly visible. Setting is $14mm$ Muller fisheye at full aperture $f/4$. The image in (a) is rendered with cos^4 factor cancelled out and usage of real aperture area during path weight evaluation. Result is a shot free of vignetting. Picture (b) includes the cos^4 term collection which apparently has only faint influence on brightness variation. Example (c) accounts for the effective aperture per pixel. Image (d) is the result of full vignetting simulation.

(a) Visible *grey* noise. A too small number of exit pupil samples causes a high deviation of the effective aperture estimation of neighbouring pixels. Additional low pass filtering on the effective aperture information could help, or just more samples.

(b) This image has been rendered with pupils of radius zero and center at the average of the positive pupil samples. Contrary to expectations, only faint artifacts at edges.

(c) This image shows the relative effective aperture area value per pixel for a pupil computation, where the same set of samples was used per pixel. New samples every pixel prevent these brightness patterns that can become visible on homogeneous surfaces.

Figure 5.15: Possible pupil sampling imaging errors caused by disadvantageous settings.



(a) at $f/4$ (b) at $f/8$ (c) at $f/16$

Figure 5.16: The lens tracing plot from Figure 5.11 repeated with the presented method. The efficiency is clearly visible when regarding the full leaving ray fan.

Conclusion

Pixel pupil sampling, for example by uniformly sampling the exit pupil to look for passing rays, has very good convergence results, even if the pupil is computed too large for example by a fair disc approximation. I repeated the lens tracing plot from Figure 5.11 with usage of the novel method in Figure 5.16. The main advantage is an optimization of eye path generation for the rendering process. The reduces variance, significantly for the first passes during rendering.

5.3.2 Progressive Rendering

The overall challenge in physically based rendering is the problem of the insufficient technique, that describes the ability of one specific method to create all possible paths. Now the Heckbert notation is of use to write some examples down shortly. The following Table 5.4 gives an overview of some combinations. Here we always assume the worst case for the surfaces, which means that S is always a singularity or Dirac pulse, like a mirror or dielectric in the scene. T is used to encompass the lens tracing from S, which is also singular of course.

By detailed examination, Table 5.4 reveals, what the last paragraph meant with *insufficient technique*, every method has problems with at least one type of path. A progressive rendering solution [Kaj86] merges several techniques to render one image. First of all, it is important to note, that all used techniques must deliver the same asymptotic result to be valid for combination.

A first, rather simple and not really progressive step, is the usage of different techniques in one pass, that are compatible in the way of path construction direction. The primary goal is to make light sources visible while using PTDL. This can be achieved by doing a PT step at scene depth $d = 0$, which means we are actually on the front surface of the lens. If we hit a light source, we can account for it. The validity can be easily evidenced by the fact that during a PTDL pass, no emission is evaluated for the position we are at the moment. We explicitly sample for incoming radiance at the location. The idea of a PT step for DL estimation can be indeed extended to all S surfaces in the scene. We use a direction sample into the singular direction, and continue the path construction with a new direction sample. For a glossy surface the directions are similar, but equal only for a mirror surface which means taking the same direction for DL estimation and continuing the path. This makes a specific handling of mirror surfaces suggestive, although they can be modelled by a Phong material description for example.

technique	path[light][scene][camera]			
	$[L][T^+E]$	$[L][D^*][T^+E]$	$[L][D^*S^+][T^+E]$	$[L][S^+D^+][T^+E]$
PT	ideal	works for large light sources	deterministic singular transport in the first is no problem to LD^*E	impossible because D in the first needs to sample a singular direction
PTDL	lens out direction singular, thus impossible to sample direct light	ideal	impossible for LSE paths, good for $LDSE$	same problem as PT
LT	the probability to hit the camera gate by direction sampling in a realistically sized scene is dramatically small and thus not considered.			
LTDL	ideal	ideal	impossible because D or light emission direction sampling must hit singularity	great results,especially for dielectric caustics

Table 5.4: A comparison of the two previously introduced rendering techniques light tracing (LT) and path tracing (PT) each also listed with a direct light estimation (DL) extension. Of course, the list of hard cases can be made even larger.

Until now we actually had no need for light tracing at all. But we should have a closer look on the first case LT^+E. In the PT approach, the direction away from the lens depends on the full path inside the lens and from the point we started. As soon as the light source is small again, we are almost lost because of the infeasible small projected solid angle covered. What seems much easier to sample, is the reverse direction. Starting samples on the light source does allow us to deterministically connect these paths with a random location on the lens. We only have to weight the path additionally by the emission character, which is apparently not singular for a normal light source. Remember that, especially a Lambert emitter has constant radiance over all angles of emission. From the just sampled point we can continue by following the refraction process to reach the sensor, provided that both points on light source and lens see each other.

The second reason, why light tracing is the better approach for DL estimation at the lens, is the secondary information formed by reflection, which was neglected before. Reintroduced only for the light tracing stage with global pupil sampling on the object side, we receive a valuable estimator to simulate reflection effects according to Section 3.2.7. Because these flare effects primarily evolve from direct light, this stage only produces paths of zero scene depth, meaning directly connecting the lens and the source. Proof of concept follows in Figure 5.17. I can only give a plot for the light tracing pass. A visual comparison with path tracing is not reasonable because the first path tracing hit on the light source for the same setting, was measured with path number 25909. For the shown plot, only 1728 paths were started at the light source.

Lens Flare Simulation

The resulting secondary information image delivers an independently viewable solution of reflection interactions. During my experiments, I observed a slightly better convergence rate, when using alternative probabilities for Fresnel event sampling inside the lens. I took the square of the transmission coefficient as probability for refraction, to allow more reflections inside the lens. Of course the weight was also changed appropriately.

The actual given results are only correct for the predetermined setting of direct light. Mirror interactions in

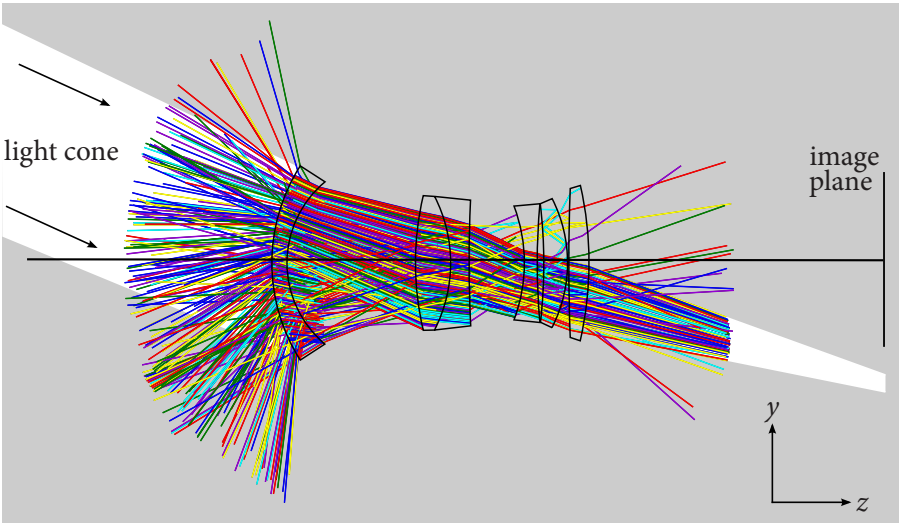


Figure 5.17: Lens tracing plot for an aperture-less Momiyama wide-angle at 36mm with fix $f/2.8$ light traced: Incoming rays were replaced by the white band, to reduce the information. The rays emanating from the front are all reflected rays, either on the first surface or later. Most of the rays miss the sensor, the white cone on the right is the image of the light source. All others, form flare effects.

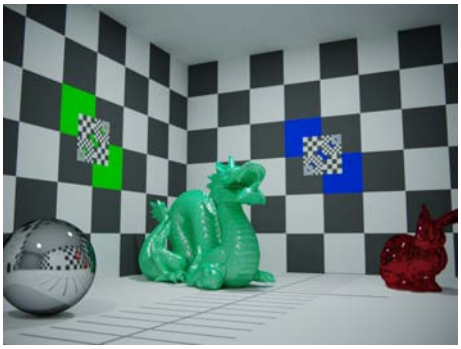
the following Figure 5.18 are not accounted for in the light tracing pass.

Conclusion

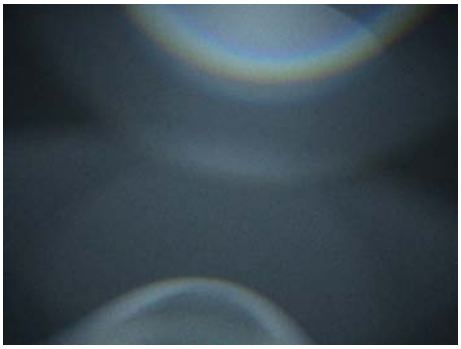
The latter mentioned flaw could be compensated by further advanced techniques for more sophisticated path generation, which could also enable us to fill the last gap in Table 5.4. By an LTDL extension, the singular LS^+DE paths for a PT become easy to create. Eric Veach also covered the topics on bidirectional path tracing (BDPT) [Vea97], the plausible next step after trying to combine LT and PT. But the way to that approach is very stony and covered by many more concerns to handle.

5.4 Remarks

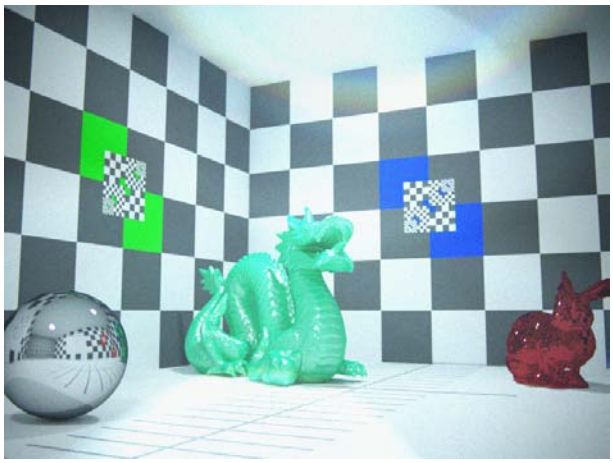
The proposed tools in the chapter allow a realisation of many more properties and effects from photography. Plane primitives, already introduced for lens surfaces can also be used in combination with wavelength specific absorption data, already known from the coating simulation. Therewith, a simulation of all thinkable filter types is realizable. The same planes can be used without any effort to define an additional gate on the object side for lens hood simulation. This resembles a further exit pupil plane, which must be passed by every path. That is nothing else but a virtual lens hood, to show the importance and positive impact on image quality, when correctly used. Flare effects can be eliminated by those mounts, allowing to photograph under extreme light conditions. But because we explicitly wanted to simulate the artifacts, I left out visual results for lens hood usage. The proposed model is able to produce paths, interacting with both lens system and scene. Reflections of light inside the lens can be simulated and thrown back into the scene, to form caustics. Finally, an extension can be added to integrate barrel flare by one further case in the main loop of the lens tracing kernel. A desired BRDF model delivers appropriate results for attenuation and new direction.



(a) Result of the modified PTDL pass



(b) Result of the LTDL pass with zero scene depth.

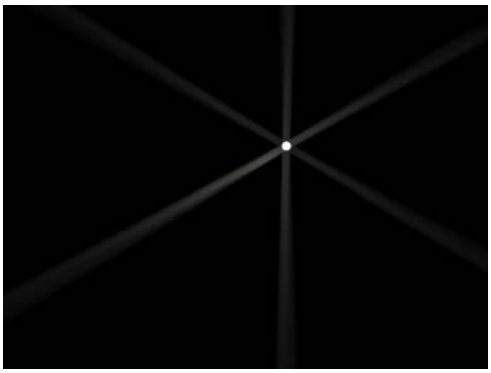


(c) Layers combined

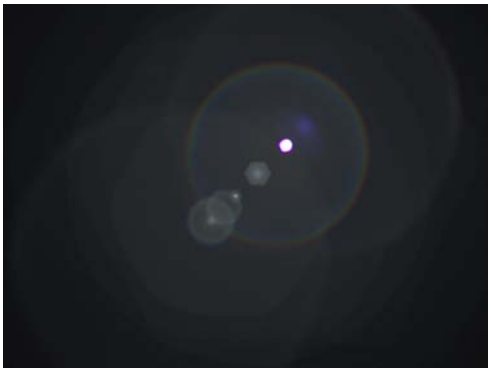
Figure 5.18: Resulting images of the above given settings from Figure 5.17.

The whole topic on sensor simulation was omitted in order not to go beyond the scope of the thesis. The type of imaging technique, either digital by a CMOS or CCD device, or analog by emulsions on film, does only play an inferior role for the proposed topics and effects.

(a) Armadillo (approx. 350K triangles) in the dark with Kimura wide-angle lens.



(b) Diffraction at the aperture simulated in a forward path tracing stage. Rendering took approximately 30 minutes.



(c) The flare image with inner lens reflection consideration, slightly brightened. With forward path tracing, this result took 130 minutes to render.

(d) Armadillo in the dark. Combined image by just summing up all previous images. Scene rendering without streaks or flares took 180 minutes.

Figure 5.19: An 800x600 sized example image stack of the proposed progressive approach with given absolute rendering times on the CPU. The machine was an 8-Core 2.5GHz Intel Xeon E5420.

6 Final Results

This chapter is a composition of additional results gained during my experiments. Most topics which are found here, are valuable results although I abstained from including them in former chapters, not to loose the central theme by too many images in between.

6.1 Aberrations

The formulated aberrations from Section 3.2.2 can be made visible with CHROMA. Unfortunately not all separately, because more knowledge from the field of optical engineering is needed to specify appropriately fitted designs like an aplanat and so on. Such systems are simply not written down, because engineers fight against them and do not waste space in books to tell how a design looks like to produce a specific effect. For the following image series I created my own little *design*, just consisting of a single biconvex lens with an object-side aperture in order to reduce marginal rays. To compute the plane of best focus, I sampled the lens with paraxial rays to find the ray with maximum height, whose focal point determines the span of spherical aberration. According to a rule given in [Ray02], the recommended focal plane lies at distance $\frac{3}{4}LSA$ away from the back lens. In order to show more details, the lower left quadrant of the results was excepted for presentation.

6.2 Astigmatism

The setting was fitted to visualise the abstract effect of astigmatism. I chose the original Petzval Portrait lens design, a famous system in the photographic history of origins. The speed of $f/3.5$ was revolutionary for the middle of the 19th century [Smi05]. I would like to give a detailed explanation to the last image of the series in Figure 6.3. To eliminate certain aberrations, different types of lenses are used, which alone only have parabola as curve, either inwards curved, towards the lens, or outwards, away from the lens surface. A compound of lenses with different shape and power results in a curved surface form of higher order, probably oscilating forth and back with the recognized results.

6.3 Lens Flares and Coatings

To show that a lens flare is a unique effect according to the given system, here some more image pairs, and also one with the proposed coating implementation. Because lens flares are only faint details under most conditions, the flare images in Figure 6.4 are shown separately. The following overlays are composed in a post process, although CHROMA could calculate both parts in one result. But for the purpose of flexibility, two images are produced.

Lens coating is, as already mentioned, an important extension to raise transmission and lower flare effects. Unfortunately due to missing coating specifications I can only give an example with a coated element of my choice. Anyway the effect is distinct, as shown in Figure 6.5.

6.4 Bokeh

Another valuable property of real lenses is the out-of-focus performance, especially for artistic purposes. The quality impression of the blur, for example how soft and constant, is called *Bokeh* [Per07, Ray02], which is primarily determined by the aperture shape and the brightness distribution of a lenses' blur circle. Two series introduced by a real example are provided in Figure 6.6 and Figure 6.7.

6.5 Further Tested Lenses

Optical engineering literature like [Smi05, Smi07] offer a wide collection of powerful lens systems to test. Examples for further, not yet used settings are given.

Pinhole Simulation

A lens without any glass element is simply a pinhole camera. To get an adequate image, one has to stop down a *bit*. Figure 6.8 confirms the statements from earlier chapters.

Telephoto Lenses

The so far not yet mentioned type of telephoto lenses are designed work at long focal lengths. A common problem is the diameter which can become large for fast versions. The shot in Figure 6.9 was done from the same position as Figure 6.4 or Figure 6.5.

Fisheye Photography

Photographs made with a fisheye lens have unique properties that stand out from the others. Depending on the design, it is possible to shoot images with 180° field of view. The field of purpose is broad, for example, environment maps can be acquired by certain transformation techniques.

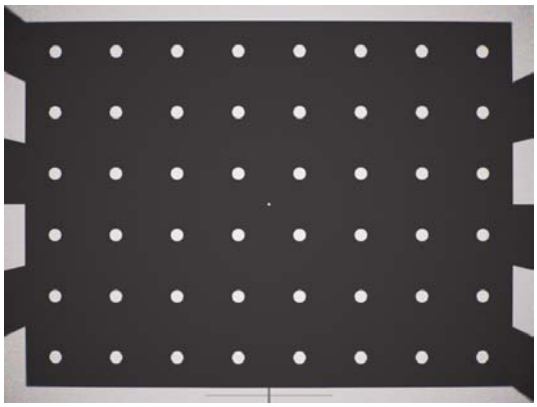
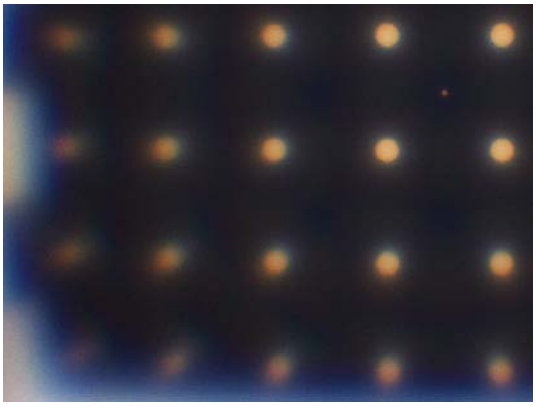
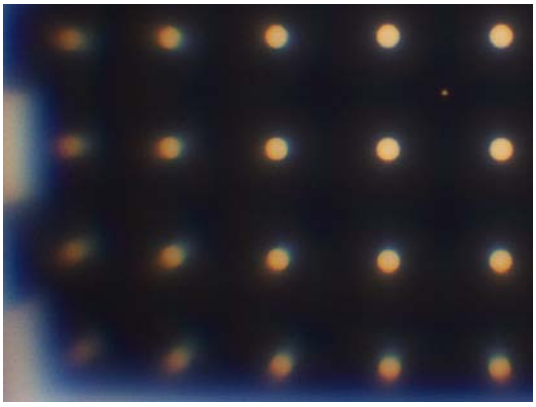


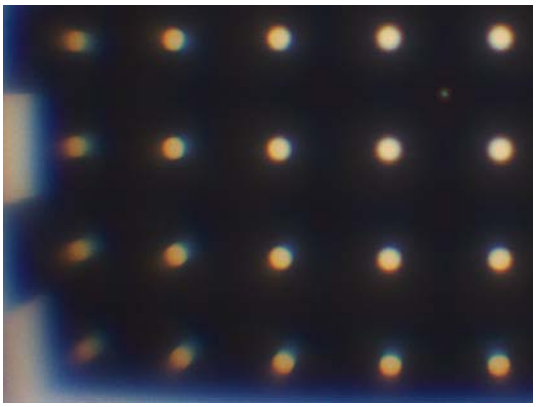
Figure 6.1: Aberration test scene overview for orientation.



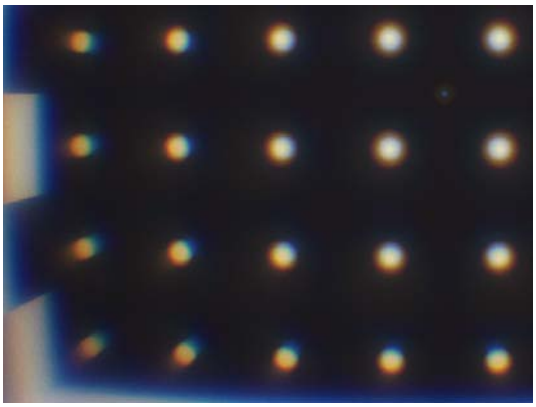
(a) Image plane at $\frac{3}{4}LSA + 1.0$mm:We are close to the paraxial focus point, as the focused center disc in the middle shows. Recognizable aberrations are barrel distortion (image looks bent), clear spherical aberration paired with field curvature (focus loss to the image margin), and coma all over the image



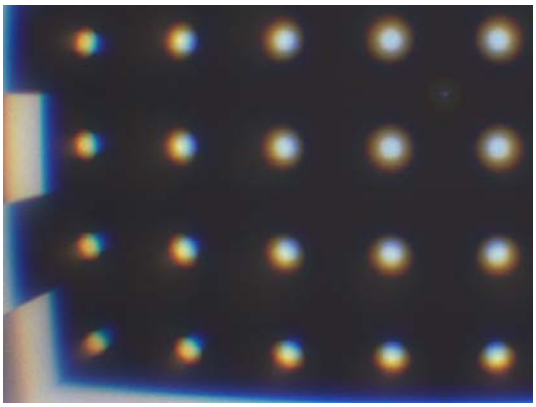
(b) Image plane at $\frac{3}{4}LSA$: The discs on the inner frame around the center disc seem focused but show first red LCA fringes,also the center disc. The second frame of discs comes into focus, although it does not look like, because of the high degree of coma in that region.



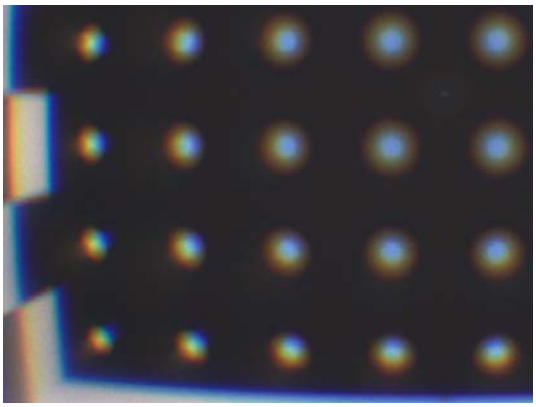
(c) Image plane at $\frac{3}{4}LSA - 1.0$mm: The separated blue part of the white central disc is clearly visible.First lateral color artifacts in the marginal region discs appear.



(d) Image plane at $\frac{3}{4}LSA - 2.0$mm: The paraxial region disc is almost gone because of the broad circle of confusion, which spread the image too wide. The side wall pattern now comes into focus.

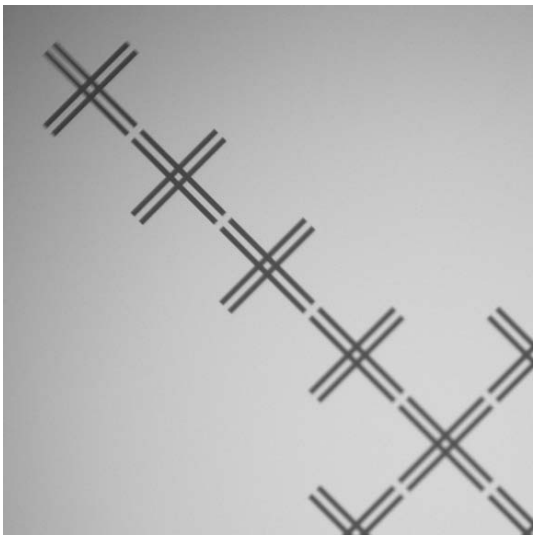


(e) Image plane at $\frac{3}{4}LSA - 3.0$mm: Now lateral color artifacts are well-defined in the marginal rind of discs. The blue part is still in focus, while the red part is not yet and shifted outbound.

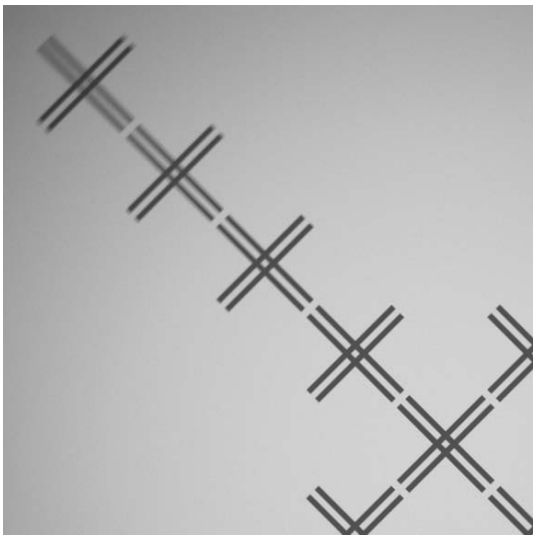


(f) Image plane at $\frac{3}{4}LSA - 4.0$mm: Image defocus now raises rapidly, and results in broadly blurred and color separated blobs.

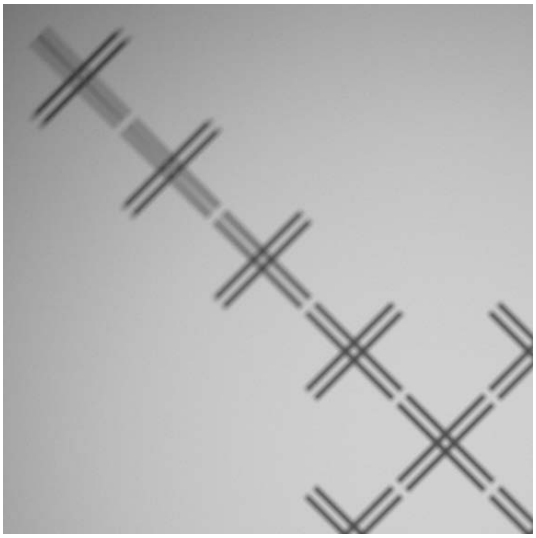
Figure 6.2: The first image is rendered with a Brendel Tessar 40mm at $f/2.8$ to give a scene overview. For the rest, the single lens with front aperture was used.



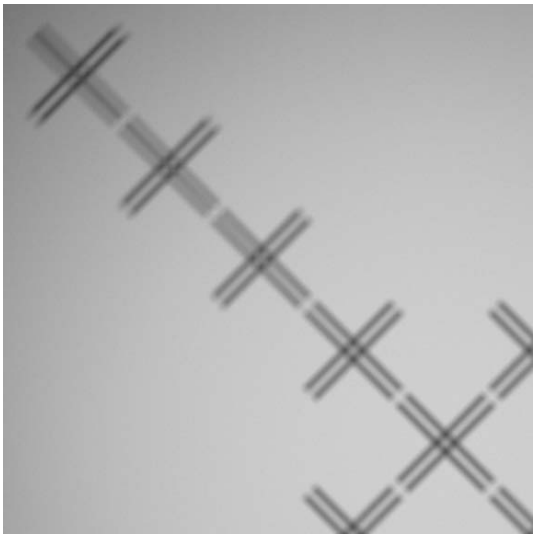
(a) Image at focal plane −1.0mm towards the lens: The marginal regions are in focus although we are *far* away from focus. A typical situation for a curved field lens, where the slope shows to the lens side.



(b) Image at focal plane: Artifacts at the border crosses. Strokes heading to the center are blurred although the perpendicular line is not. That is the practical evidence for different focal planes for sagittal oriented scene structures (towards the image center) and transversal oriented content (tangential to the circle around the image center.)

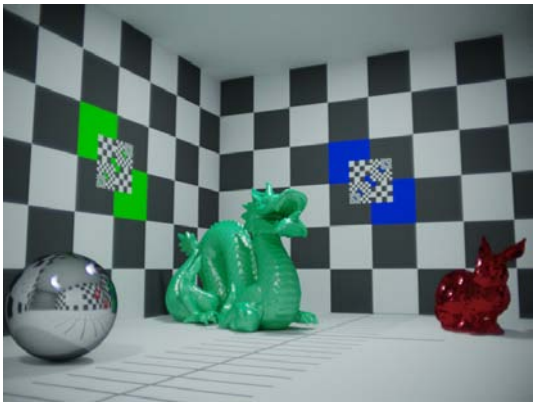


(c) Image at focal plane +1.0mm: This image underlines the observation. The field is not only curved in one direction. It seems, like the tangential strokes of the out-most crosses are still more in focus than the ones nearer to the center.

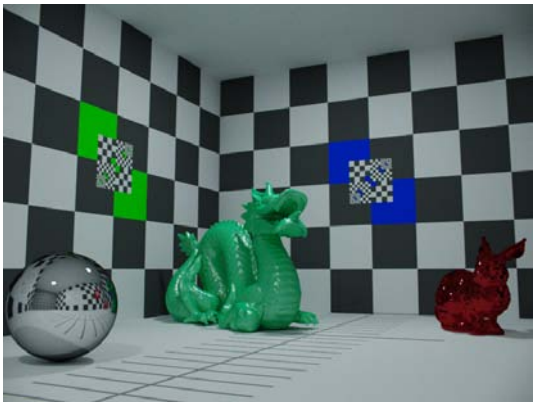


(d) Image at focal plane +1.5mm: Image is overall blurred except the outmost tangential structures, which seem still more sharp.

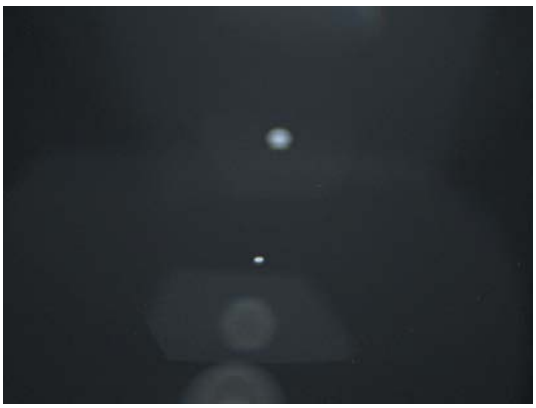
Figure 6.3: Astigmatism rendering series



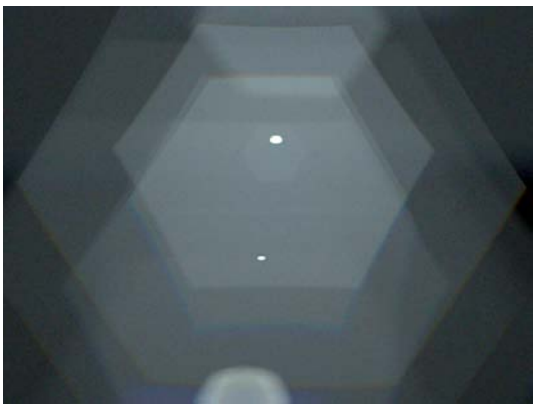
(a) CHROMA test scene at $f/2$. By the way, the image is a nice example for apparent aberration in a practical design. The lens suffers a high amount of coma, visible at the light source reflections in the mirror ball and on the glass bunny.



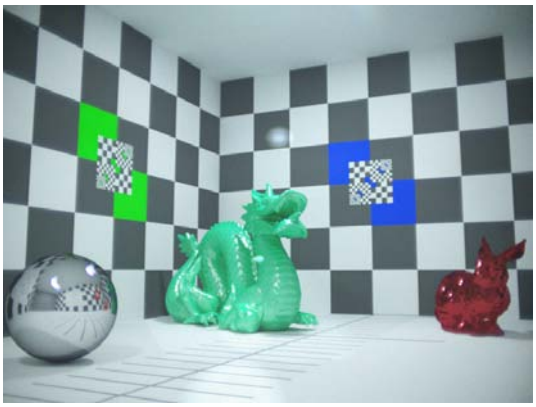
(b) CHROMA test scene at $f/5.6$. The just mentioned coma is gone by stopping up.



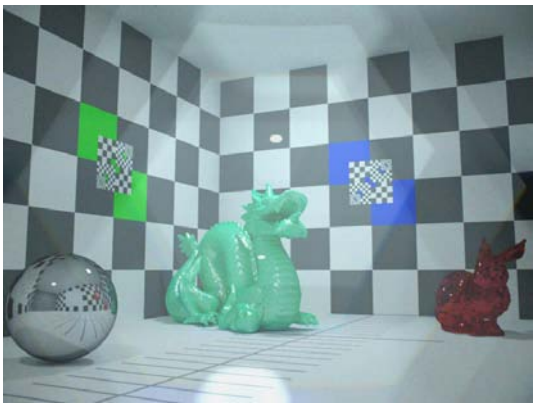
(c) The flare produced by the lens appears as considerable veiling glare, which destroys the scene contrast. The aperture ghosts are faint.



(d) In the stopped up version, the veiling glare begins to form aperture ghosts in the image center. Both spots concentrated to visible ghosts of the light source.

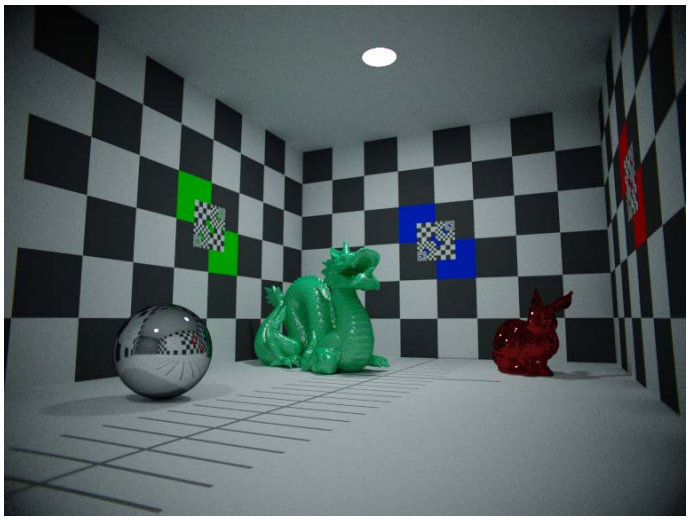


(e) The image suffers an overall veiling glare.



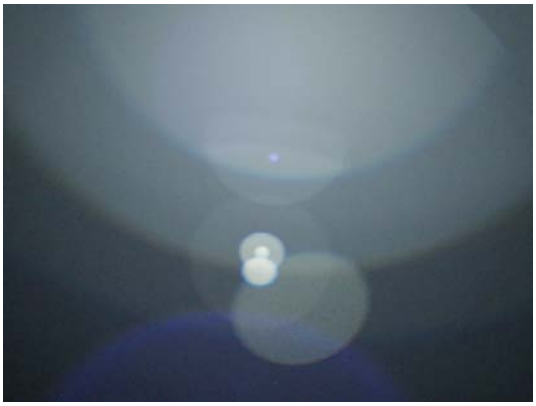
(f) Lost contrast in this version concentrates on the center of the image.

Figure 6.4: The setting is a Kimura lens at 36mm focal length.

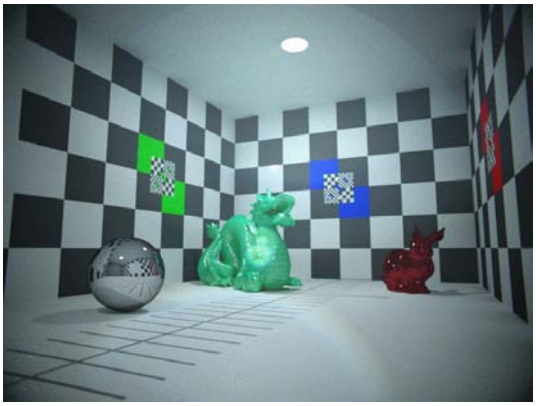


(a) CHROMA test scene: The high amount of optical vignetting contributes a nice frame to the scene.

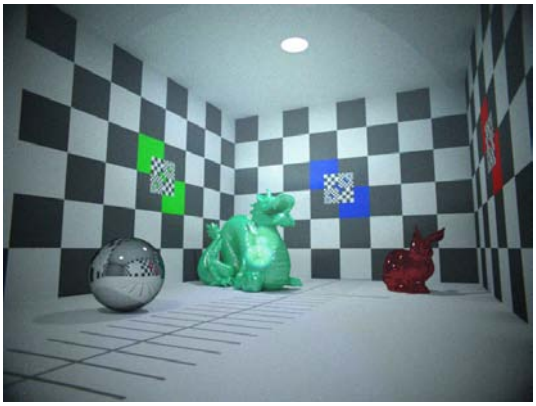
(b) As the lens has no aperture, all flares are round shaped. The high variance region in the lower middle should result in a similar circle like the large rainbow around, but this one seems to have horrible paths.



(c) The second element in the lens was coated from both sides and gives this result. Although not visible here, the RGB average image brightness increased by (+8.8%, +10.0%, +3.8%).



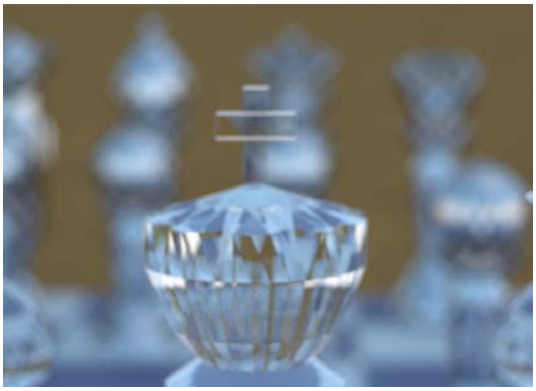
(d) In the compound image, most flares vanish on the bright scene background.



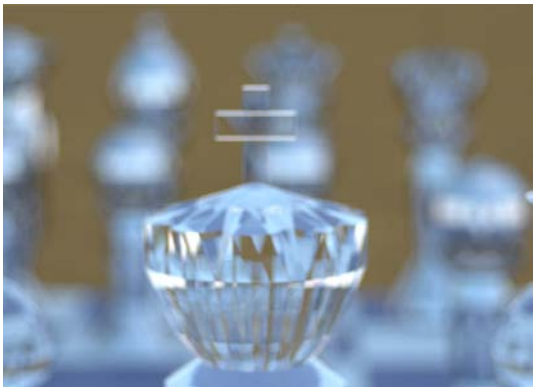
(e) The version with coated element seems to have a higher amount of veiling glare, explainable by the increased transmission.

Figure 6.5: The setting is an aperture-less Mori wideangle lens at 25mm focal length with $f/2.8$ [Smi05].

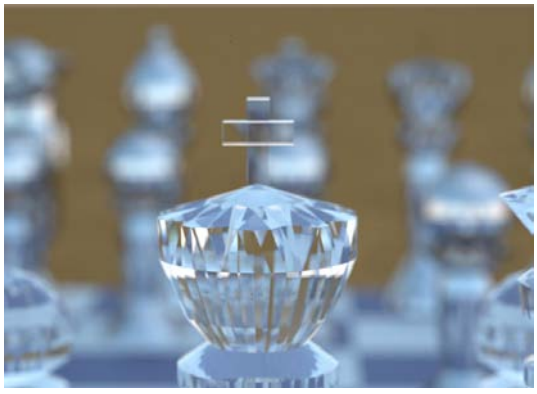
(a) Photograph taken with a six bladed lens diaphragm at full aperture. The contrast was reduced to clarify the blur disks of the point like reflections in the background. The aperture shape is recognizable and the blur is almost equal except a small band near the margin.



(b) Chessboard scene shot, with a four blade aperture forming a diamond shape. The out of focus region looks turbulent and patterned because of the diagonal strokes caused by the bokeh.

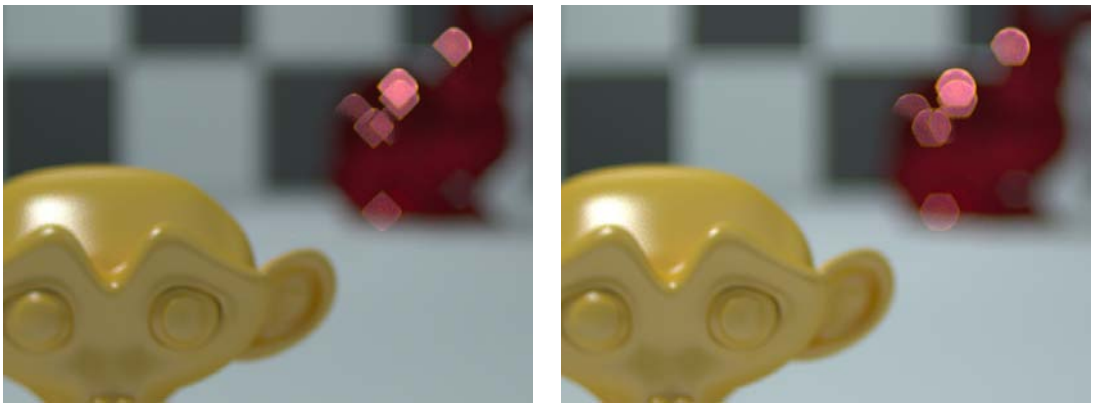


(c) The same setting, but now with a six bladed diaphragm. The out of focus region look much smoother and more well formed.



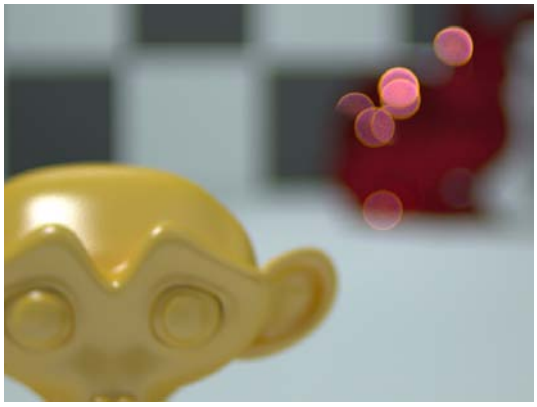
(d) For comparison, I used the implemented aperture model with a thin lens approximation and equivalent settings. Typical for it, the image looks artificial, and the Bokeh with its perfect equal distribution over the discs just as well.

Figure 6.6: Setting is a a 80mm Brendel Tessar lens [Smi05] at full aperture $f/2.8$ closeup to the King, to have minimal depth of focus.



(a) Four blades

(b) Six blades



(c) Twelve blades

Figure 6.7: The CHROMA test scene has point like reflections in the scene which form non-overlapping clearly visible Bokeh pattern. In picture (c) also a smoothing enhancement of the blurred wall pattern seem to be visible, in contrast to the other two images. The used lens was Rosier Double Gauss lens [Smi05] with 56mm focal length at $f/1$.

Figure 6.8: Pinhole rendered version of the CHROMA test scene, this time with a CIE-A light source. Focal length is 36mm, with 0.18mm aperture radius or $f/100$ respectively. The aperture shape is nicely visible at the highlights on the glass bunny.

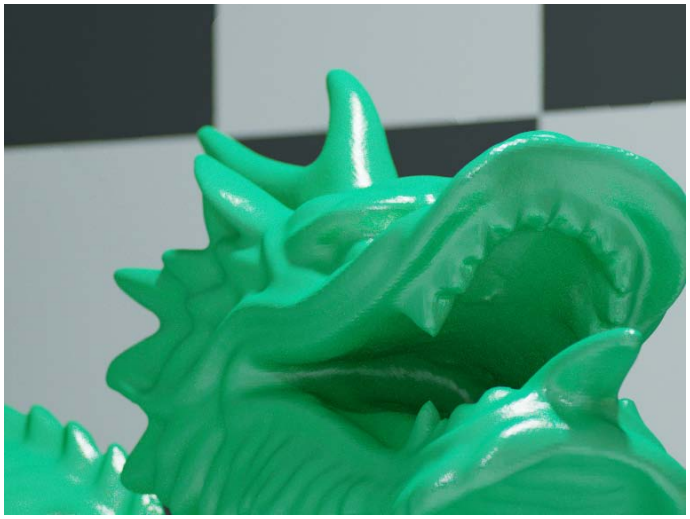


Figure 6.9: CHROMA test scene shot, done with a Kingslake design at 225mm, $f/8$.

(a) History Museum scene taken with a 12mm Muller fisheye at $f/5.6$ and 144° field of view.

(b) The same scene, but now with a 179° 10mm Mitsuaki design at $f/4$. from a central position that leads to an impressive environment map. The used daylight model [AJP99] contributes over 14 stops of dynamic range.

Figure 6.10: Fisheye lens shots.

Figure 6.11: An enlargement of the left window in Figure 6.10a reveals the faint lateral color shift, which was the eponm of my renderer CHROMA.

7 Summary and Future Work

During this thesis, I gave an insight to various topics concerning the field of optical engineering. The results showed that a simulation of all aberrations is absolutely possible, allowing us to introduce these effects to a rendering approach. The virtual view through a lens design becomes possible. Photo-realistic rendering through mounts like fisheye lenses yield a unique look, and are above all compatible to existing work-flows, which handle these shots for panoramic transformations. We are able to use optical devices at any reasonable or unreasonable setting in the simulation and can study its characteristics. Next step could be a comparison of achieved results with performance analysis output from lens design software like *OSLO*. Concrete applications in the domain of digital image forensics have to be considered in future. The synthesis of lens flares based on correct lens data provides information about correct shape, position and contrast influence.

The rudiment introduction of wave effects by the geometrical theory of diffraction definitely has to be extended by the related topic of interference. First implementations based on ray length consideration did not give meaningful numbers because of high deviations caused by floating point imprecision. Aveneau and Mériaux also made continuing proposals on this topic in [ABM00]. Primary aim should be simulating the formation of well known and analytically verifiable diffraction pattern like the Airy Disk [BW75, Smi07]. Further investigations are required to prove whether the proposed method suffices the needs of an image synthesis scenario at all.

The final intention of this thesis is to summarize the possibilities of real lens simulation with consideration of all possible paths directly adoptable from Fermat's principle. This case was introduced by a progressive rendering approach, to allow for optimisations and specialised methods in the different passes. The method of sampling the pixel specific pupil for path generation at the sensor enabled us to have a stable rate of passing rays while sampling, with increasing f-stop value. Suggestions concerning similar ideas for a light tracing approach have to be considered in future. Merged with techniques for direction generation on an environment map [DH08], we should receive an equivalent approach, providing flare effects from daylight models for example. A further idea to mention is an equivalent possibility of pupil sampling for light tracing. The proposed pixel pupils could be evaluated for the field of view region with a certain discretization of the domain. The direction from a scene point to the lens is suggested to index into a map similar to an environment map and can provide the introduced pupil sampling technique for the light tracing stage.

Finally, to mention the next step in progress, the whole lens simulation, based on the small set of mathematical primitives will be ported to a GPGPU environment, where the graphics card computes the lens tracing part in parallel to the CPU. This follows the inclusion of the hereby presented results without extra cost on the CPU. Another new aspect will be the precomputation of the whole radiance transmission in a huge *transmission map*, enabling us to approximate the complete image formation with one lookup and possible interpolations on the map data.

Bibliography

[ABM00] Lilian Aveneau, Phillippe Blasi, and Michel Mériaux. The wave nature of light in monte-carlo rendering. In *Proceedings of the 16th Spring Conference on Computer Graphics*, 2000.

[AJP99] Brian Smits A. J. Preetham, Peter Shirley. A practical analytic model for daylight. In *SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 91–100. ACM Press/Addison-Wesley Publishing Co., 1999.

[AM99] Lilian Aveneau and Michel Mériaux. Rendering polygonal scenes with diffraction account. In *Seventh International Conference in Central Europe on Computer Graphics and Visualization (Winter School on Computer Graphics*, 1999.

[BSMM05] I. N. Bronstein, K. A. Semendjajew, G. Musiol, and H. Muehlig. *Taschenbuch der Mathematik*. Verlag Harri Deutsch, 6th edition, 2005.

[BW75] Max Born and Emil Wolf. *Principles of Optics*. Elsevier, 1975.

[CIE04] CIE - Commision Internationale de l'Eclairage. *Selected colorimetric tables*, 2004. `http://www.cie.co.at/publ/abst/datatables15_2004/CIE_sel_colorimetric_tables.xls` [accessed 2009, February 14th].

[DH08] Holger Dammertz and Johannes Hanika. Plane sampling for light paths from the environment map, 2008. Submitted to Journal of Graphics Tools.

[Dut03] Philip Dutré. Global illumination compendium. Technical report, Department of Computer Science, K.U.Leuven, 2003.

[Kaj86] James T. Kajiya. The rendering equation. In *SIGGRAPH '86: Proceedings of the 13th annual conference on Computer graphics and interactive techniques*. ACM Press, 1986.

[Kel62] Joseph B. Keller. Geometrical theory of diffraction. *J. Opt. Soc. Am.*, 52(2):116–130, 1962.

[KMH95] Craig Kolb, Don Mitchell, and Pat Hanrahan. A realistic camera model for computer graphics. In *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 317–324. ACM, 1995.

[LW94] Eric P. Lafortune and Yves D. Willems. Using the modified phong reflectance model for physically based rendering. Technical report, Department of Computer Science, K.U.Leuven, 1994.

[MPM90] Derek A. McNamara, Carl W. I. Pistorius, and J. A. G. Malherbe. *Introduction to the uniform Geometrical Theory of Diffraction*. Artech House, 1990.

[MT97] Tomas A. Moeller and Ben Trumbore. Fast, minimum storage ray-triangle intersection. *Journal of Graphics Tools*, 2, 1997.

[Per07] Michael R. Peres. *Focal Encyclopedia of Photography*. Focal Press, 4th edition, 2007.

[PH04] Matt Pharr and Greg Humphreys. *Physically Based Rendering: From Theory to Implementation*. Morgan Kaufmann, 2nd edition, 2004.

[PPBS05] Frank L. Pedrotti, Leno S. Pedrotti, Werner Bausch, and Hartmut Schmitt. *Optik fuer Ingenieure: Grundlagen*. Springer, 2005.

[Ray02] Sidney F. Ray. *Applied Photographic Optics*. Focal Press, 3rd edition, 2002.

[Rob08] Barbara Robertson. Making of walle: "vom abfall zum schatz". *Digital Production*, (5):156–162, 2008.

[Sch09] Schott AG. *Optical Glass - Data Sheets*, 2009. `http://www.schott.com/advanced_optics/english/our_products/materials/data_tools/index.html` [accessed 2009, March 22nd].

[Shi05] Peter Shirley. *Fundamentals of Computer Graphics*. A K Peters, Ltd., 2nd edition, 2005.

[Smi99] Brian Smits. An rgb-to-spectrum conversion for reflectances. *Journal of Graphics Tools*, 4(4), 1999.

[Smi05] Warren J. Smith. *Modern Lens Design*. McGraw-Hill, 2nd edition, 2005.

[Smi07] Warren J. Smith. *Modern Optical Engineering*. McGraw-Hill Professional, 4th edition, 2007.

[Vea97] Eric Veach. *Robust Monte Carlo Methods for Light Transport Simulation*. PhD thesis, Stanford University, 1997.

[Wat02] Alan Watt. *3D-Computergrafik*. Pearson Studium., 3rd edition, 2002.