

UTRECHT UNIVERSITY

MASTER THESIS PROJECT

Hair Rendering: Importance Sampling of Dual Scattering Approximation

Author:
Jeffrey LEMEIN

Supervisor:
Prof. Remco VELTKAMP

April 13, 2020

Abstract

Rendering human hair models using individual hair fibers is a challenging task. It is challenging because of the extensive amount of hair fibers required to render a realistic model. Moreover hair fibers are very thin, resulting in noise in the renderings. To reduce this noise in a physically accurate way, more samples need to be taken. Marschner et al. (2003) proposed a physically based single fiber scattering model to render realistic human hair fibers. It is considered the founding work on current approach to physically based rendering human hair fibers.

When rendering light colored hair, multiple fiber scattering is essential for the appearance of the hair color. Zinke et al. (2008) extended the work of Marschner by splitting multiple scattering up into two components: global multiple scattering and local multiple scattering. Global multiple scattering approximates the multiple scattering contribution, thereby reducing the rendering time considerably. Local multiple scattering resembles the Marschner model to keep the single fiber scattering characteristics. This approach is known as the dual-scattering method.

Importance sampling is a widely used noise-reduction technique to speed up renderings. By applying importance sampling, samples are not taken randomly (as is the case for uniform sampling), but are sampled according to a probability density distribution that favors samples that contribute more to the output rendering. This reduces noise faster and thus reduces the amount of samples that are required to produce noise-free renderings. Both uniform and importance sampling should eventually converge to the same result, with importance sampling reaching it faster.

d'Eon et al. (2013) proposed an importance sampling strategy that works particularly well for the Marschner model. This importance sampling strategy can also be applied to the dual scattering method. It is therefore very interesting to evaluate whether applying importance sampling to the dual-scattering method reduces the noise as well. This comparison forms the contribution of this report.

This report concludes by measurement and visual inspection that importance sampling for the dual-scattering method does not significantly increase the quality of the renderings compared to uniform sampling. Such a small increase in quality is not what was expected. A possible explanation is that the effect of multiple scattering is less dependent on specific sampling directions. It essentially smoothens the contribution from any direction and therefore, importance sampling does not significantly outperform uniform sampling.

Contents

Chapter 1

Introduction

Hair rendering has always been a challenging task due to the amount of hair strands and the complex scattering behaviour between them. A human head may consist of over 100.000 hair strands and rendering them in a realistic and efficient way is challenging. Furry animals may even consist of over millions of hair strands. It is easy to see that the amount of strands provides us with a big challenge. Efficient rendering models should be devised.

Hair rendering becomes more important as more physically accurate models are constructed. The following number of industries benefit from realistic looking hair models:

- Animation movie industry: to efficiently render realistic hair, instead of simplified models.
- Game industry to enhance realism and for better visual effects.
- Clothes manufacturers: to be able to realistically render custom fabrics and see what the visual appearance of the fiber structure will look like under different lighting conditions.
- Hair styling industry: to realistically render the appearance of hair styling products applied to the hair.

Since rendering individual fibers is very time consuming, a lot of work has been put into approximating the appearance of a hair volume. For example, by applying transparent textures to a bounding volume to give the illusion of a realistically looking hair model. This approach may work well for a still image or for applications that do not focus on hair so much, but is not very flexible for movie or game industries. These industries often have dynamic environments in which hair might be animated or lights may change position, leading to shifting highlights that are more difficult to manage using non-physical based approaches.

Nowadays, improved hardware makes it possible to render models of individual fibers. This opens the door to physically based models based on ray-tracing. Path-tracing is the most realistic way of rendering the fibers by considering all the scattering events between light rays and individual fibers and consequently adding up the contributions along the way. It still takes up a substantial amount of time to render a single frame. The time is heavily dependent on the amount of hair fibers, but to overcome the hundred of thousands of hair strands, efficient rendering models should be devised.

1.1 Goal and approach

The work in this thesis is primarily concerned around the work of three publications. It starts with the Marschner model proposed by Marschner et al. [?] which forms the foundation of hair rendering in this thesis. The dual scattering model by Zinke et al. [?] extends the Marschner model with an efficient strategy to render volumes of hair. The third work is the importance sampling strategy devised by d'Eon et al. [?].

Goals

In this thesis project the goal is to extend the dual scattering model from Zinke et al. [?] with the importance sampling strategy proposed by d'Eon et al. [?]. Importance sampling is a variance reduction technique that works very well together with Monte-Carlo integration. The focus will be on the importance sampling strategy. There are in general two problems that need to be tackled.

- 1 The method in this report is based on the traditional Marschner model, which is not strictly energy conserving. The improved approach by d'Eon et al. [?] fixes the shortcomings and they also provided an importance sampling method in their follow up paper [?]. This importance sampling strategy is used in this report to importance sample the dual scattering method. Since the dual scattering method is based on the "flawed" Marschner model, this can lead to a slightly darker or lighter image. The goal is to see whether the results are still acceptable.
- 2 A second problem is the global illumination computation, given a direction through the hair volume. There are different ways to gather this and in this report a voxel grid is used to store the density of the hair model at specific locations in three-dimensional space. This provides some issues with aliasing at edges of voxel cells, that need to be solved.

The ultimate goal is to see whether the importance sampling method proposed by d'Eon et al. [?] can be used efficiently for the dual scattering

approximation model by Zinke et al. [?] and thereby speed up the rendering time.

Approach

To evaluate the importance sampling strategy, rendering results are generated to visually compare results obtained by using uniform sampling with the results obtained by using importance sampling. This gives a gut feeling about the efficiency of the sampling strategy. Additionally I will vary the number of samples per pixel to compare the same output renderings, with the only difference in increasing numbers of samples per pixel. This evaluation is done by computing the variance between these images.

Moreover, the variance of the produced renderings are compared with the ground truth. Setting the ground truth to a real-world hair model is not practical, because a real world hair model is too complex to accurately model in modeling software. Think about the hair orientations, pigment changes along the hair strand and hair cut in general. Instead, the ground truth is set to the results obtained by using 512 samples with uniform sampling.

The rest of this report is divided into a background section, where the basics are explained regarding radiometry and hair rendering. Related work explains the different methods that exist to render hair. Here the Marschner model and the Dual Scattering approximation model are explained in depth, because these models are the basis for the work in this thesis. Additionally, an importance sampling strategy for physically based hair fiber models is explained from d’Eon at al. [?]. After the related work, the approach will be explained on how to compare renderings from uniform sampling and importance sampling by looking at the variance. After that, the results are presented showing the efficiency and the analysis of the importance sampling function versus the uniform sampling version of the dual scattering approximation model.

Chapter 2

Background

Hair comes in different styles. According to Ward et al. [?] hair can be smooth, jagged, wavy or curly depending on the ethnic group. Three different ethnic groups are distinguished: Asian, African and Caucasian hair. A hair fiber can either have a circular or elliptical cross section. Asian hair tends to be very smooth and regular with a circular cross section, whereas African hair is very irregular and has an elliptical cross section. Caucasian hair is in between and can be a mixture of both properties.

The follicle is the active part of hair under the skin and produces keratin proteins that compose the hair material [?]. The visible part of the hair is called the hair shaft. Hair has an almost uniform cross section, natural twist and natural curvatures all along [?]. It makes the hair appear curly, straight, fuzzy, smooth, etcetera. These parameters are characteristic for the ethnic group.

A hair fiber of the human scalp consist of three layers. The outside layer is the cuticle. The cuticle is a thin sheath of protective layers that surround the inner cortex. It forms the interface between the fiber and the air. The protective sheath consist of flat cells on top of each other. Because of the overlapped arrangement the surface normal is tilted slightly from the overall normal of the fiber's surface. According to Marschner et al. [?] the surface normals are tiled by approximately 3 degrees toward the root. See figure ?? for a view on the cuticle scales.

The center of the hair fiber consists of the medulla: a pigmented core. The pigments give hair its color. The cortex is the core of the fiber and provides the strength for the hair fiber. The cortex is filled with keratin, contributing 90% of the total weight [?]. Keratin is a stiff material that causes hair to be easily bend and twisted, but hard to shear and stretch.

Hair consists of amorphous proteins that act as a transparent medium with an index of refraction of $\eta = 1.55$ [?]. A hair model will later be introduced that models hair fibers as dielectric cylinders where the index of refraction is important in deciding the refraction angle and the amount of

Fresnel reflection.

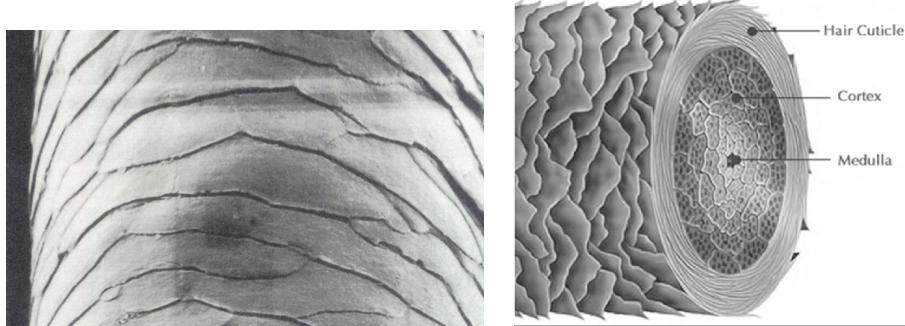


Figure 2.1: On the left the tilted cuticle scales on the outside of the hair fiber (taken from Marschner et al. [?]). On the right a cross section overview of the core of a hair fiber, consisting of the cuticle scales with the cortex and medulla (taken from Morganti [?]).

2.1 Rendering hair

Marschner et al. [?] explain that hair rendering is a complex task, because local and global properties of hair must be taken into account. Local properties describe how light interacts with hair fibers and global properties describe the propagation of light through a volume of hair. In general, hair can be modelled explicitly or implicitly.

2.1.1 Explicit Representation

Hair rendering research began by analyzing hair models using an explicit representation. This means that a hair model is represented as individual strands, or one-dimensional curves in three-dimensional space. Building on these foundations, researchers tried to find an efficient model to render a full head of human hair. Though several paths have been followed, modelling a full head of hair remains a challenge due to the geometric complexity and the thin nature of an individual strand coupled with the complex scattering effects and shadows that occur among the hairs [?].

2.1.2 Implicit Representation

Implicit hair representations stay away from the complexities of individual strands. With implicit representations, hair is rendered by approximating the appearance of the several strands using simpler and more efficient rendering strategies. Voxel grids, multi-layer textures or polygonal boundary

representations are some of the possibilities to efficiently render a full hair model.

Implicit hair is usually modelled as a volume or polygonal boundary representation. Implicit representations work great when looking at hair from a distance, because at a distance the detail is hardly visible. Moreover, it is much faster in rendering and still giving good-looking results.

The downside of implicit representations is that it is not physically based and it is not flexible. Scattering behavior are usually simplified to work only for specific viewing directions. For physically based rendering, the explicit representation is the only way to go.

2.1.3 Geometric complexity

?? A human scalp may consist of over 200.000 hair strands. Some animals, such as bears might realistically be rendered with millions of hair strands. It is clear that the amount of hair fibers lead to a geometric complex challenge.

A hair fiber is naturally represented as a curved cylinder. For rendering, these cylinders are usually approximated to make computing intersections more efficient. There are strategies to render curves. Curves can be treated as:

- Connected triangle strips facing the camera. Triangles are easy to intersect and a strip that always faces the camera will always be intersectable.
- Cylindrical primitive. This is the most realistic solution, because it comes closest to how a hair fiber is physically represented. It is similar to the previous approach, except that the normal is now adjusted to account for a cylindrical shape.
- Trigonal prisms. This is an extension to camera facing triangle strips, making three ribbons and attaching the ends together to form a trigonal prism.
- Ribbons, in which a triangle strip is constructed following the curvature of the three-dimensional curve. This triangle strip is not necessarily camera facing.

In this thesis the Physically Based Ray-Tracing (PBRT) framework is used. This is an open source physics based rendering framework created by Pharr et al. [?]. It is used by students to understand the theory of physically based rendering and academics to test their physics based rendering algorithms. The framework is extensively documented in their corresponding book [?].

In PBRT hair fibers are represented as a three-dimensional curve with a certain defined width. When rendering, these curves are treated as cylinders

by creating a camera facing triangle strip where the normal is adjusted to reflect a cylindrical primitive (the second option from the list above).

2.1.4 Aliasing problem

A hair fiber is extremely thin. So thin that it is much smaller than the size of a pixel. When tracing a ray through a pixel it will most likely miss the hair fiber. This results in aliasing problems. According to Hadap et al. [?] there are in general two strategies to perform anti-aliasing:

- **Oversampling:** With oversampling you cast a lot of rays through a single pixel so that enough rays hit the fiber to smooth out the aliasing effect. This is a very expensive solution, because it increases rendering time considerably. For example, tracing 128 samples through a pixel could lead to a 128 times higher rendering time.
- **Pixel Blending:** blend the contribution of each hair fiber per pixel. This comes close to rasterization. Assuming that a curve is represented as a one-dimensional curve in three-dimensional space, then a fiber consists of a sequence of curve points. Drawing a spline based on these curve points, one can compute the contribution at each fragment of the curve, then backprojecting on the camera plane to find out which pixel is affected and add the contribution to the pixel. Blending needs to be performed to incorporate fibers that are positioned behind each other.

Physically based rendering is based on ray-tracing, so in this project the aliasing problem is overcome by oversampling and by giving each hair fiber a certain width when rendering. Increasing the width of a hair fiber increases the chance of being intersected. This makes it easier to ray-trace, thereby reducing the need for many samples. Different software packages have different optimizations to render curves.

2.1.5 Mathematical Notation

The notation that is used in this thesis is the same as the one used by Marschner et al. [?] and is common in the field of hair rendering. A hair strand is represented as a one-dimensional curve in three-dimensional space. A uvw coordinate axis is attached to any point along the curve.

The u -axis is tangent to the hair fiber and pointing in the direction from the root toward the tip. v and w complete a right-handed orthonormal basis and are situated in the normal plane of the hair fiber. If the cross section is elliptical, then v represents the major axis and w the minor axis. The direction where the illumination is coming from is denoted by ω_i and the direction in which light is scattered is ω_r . These directions are expressed in spherical angles (θ_i, ϕ_i) and (θ_r, ϕ_r) respectively.

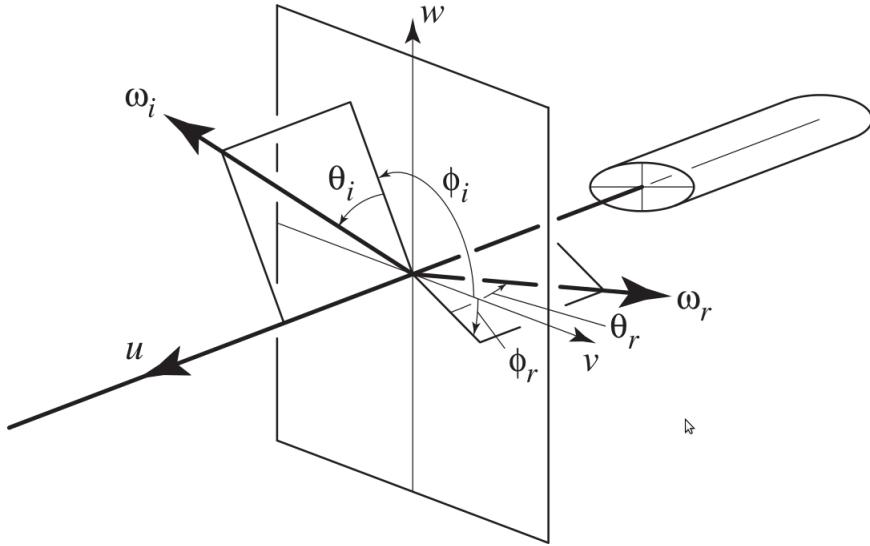


Figure 2.2: An overview of the coordinate axes and the spherical angles with respect to a hair fiber. This is the convention used in this thesis and is used by Marschner et al. and most other hair rendering algorithms. Picture taken from Marschner et al. [?].

The longitudinal inclinations with respect to the normal plane are denoted θ_i and θ_r and are measured so that 0° is perpendicular to the hair and 90° is u and -90° is $-u$. The azimuths around the hair are denoted ϕ_i and ϕ_r , measured so that v is 0° and w is $+90^\circ$.

Derived angles are used, such as the difference angle $\theta_d = (\theta_r - \theta_i)/2$. The relative azimuth $\phi = \phi_r - \phi_i$. The half-angles are $\theta_h = (\theta_i + \theta_r)/2$ and $\phi_h = (\phi_i + \phi_r)/2$. Figure ?? shows an overview of the axes and spherical angles.

2.2 Radiometry

It is important to understand the terminology in radiometry to get a good understanding of the theory that is explained in this thesis. This section explains the radiometric quantities that are important to understand for hair rendering and rendering in general.

2.2.1 Power and radiant flux

Power is the rate at which energy is transferred and is expressed in watts (W). A single watt describes the transfer of a single Joule (J) of energy per

unit time (s). All electronical devices consume power in one way or another. In terms of rendering we most often talk about power in the form of light propagating through the scene.

Lights are emitting light particles through the scene at a certain rate. These elementary light particles are called photons. The stronger the light, the faster a light source emits photons in the scene. Photons exhibit a wave-particle duality. On one side they can be treated as minuscule particles and on the other side they can be treated as waves, or more specifically: electromagnetic waves.

The rate at which a light source sends out electromagnetic waves is called radiant flux. Radiant flux Φ is the power of electromagnetic waves and describes the transfer of radiant energy (J) per unit time (s). Radiant flux is therefore also expressed in watts (W).

2.2.2 Solid Angle and Radian Intensity

A two-dimensional angle can be represented in polar coordinate space by drawing two lines from the center of a circle outward. The angle is then represented as the difference in orientation between the lines (see figure ??). If an object can be drawn so that the angle (if extending the lines) can fully encapsulate the object, then the object is considered to be subtended by the 2D angle. Angles are represented as degrees or radians, but radians are more convenient for mathematical purposes.

A solid angle Ω is the equivalent to a two-dimensional angle, but solid angles work in three-dimensional space. It therefore can be visualized as drawing a cone outward from the center of a sphere. Similarly to the two-dimensional case, a three-dimensional object can be subtended by a solid angle. Note that objects of different sizes can have the same solid angle. This is possible as the bigger object is placed farther away from the center point of the sphere (see figure ??).

A solid angle is a dimensionless unit of measurement called a steradian (sr). Steradian is coming from squared radian as is an obvious generalization when you realize angles are represented in radians.

The radiant intensity I is a measure of the intensity of electromagnetic radiation. It can be seen as the energy flow divided by the solid angle in which it is transmitted. Radiant intensity is defined as radiant power per unit solid angle ($W sr^{-1}$).

2.2.3 Irradiance

Formally, irradiance E is defined as the amount of power of electromagnetic radiation per unit area incident to a surface. The unit for irradiance is watts

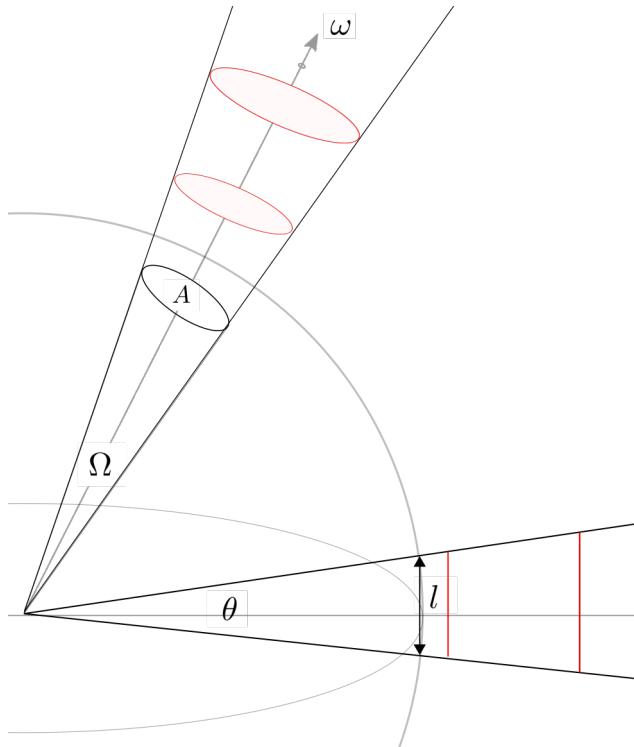


Figure 2.3: Graphical representation of angles and solid angles. θ represents a two-dimensional angle. If the circle is considered a unit sphere, then the length of line l equals θ radians and it subtends the red lines on the right. The solid angle Ω is a three-dimensional angle represented by a cone. At the unit sphere, the solid angle represents a surface area A . The red disks shown on the outside represent three-dimensional objects that are subtended by the solid angle Ω . Larger objects can be subtended by the same solid angle as they are placed farther away from the center of the sphere. As the solid angle becomes infinitely small, we obtain a direction vector ω .

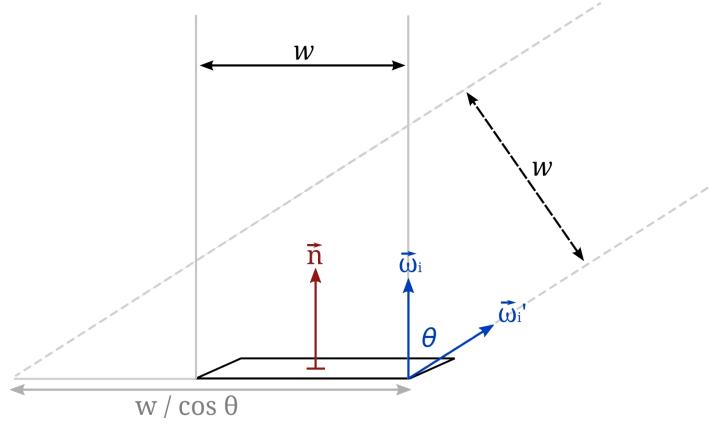


Figure 2.4: As the incoming light hits the surface at a larger angle, the irradiance decreases because the same amount of light will be spread over a greater surface area.

per square meter ($W \cdot m^{-2}$). If the light is shining directly down on the surface area, the irradiance can be described as:

$$E = \frac{\Phi}{A} \quad (2.1)$$

Figure ?? shows a simplified scenario of a directional light shining on a surface. Irradiance is the amount of radiant flux (or power) that is received per unit area. If we consider a rectangular bundle of light with width w that hits the surface from a perpendicular direction ω_i , then it is obvious to see that all the light falls into an equally sized rectangle on the surface (equation ??). The irradiance would be equal to the power (or wattage) of this bundle of light, divided by the surface area.

As the angle with the surface normal \vec{n} increases, the direction from which it hits the surface ω'_i becomes smaller. As a result, the bundle of light spreads its power over a larger area and therefore the irradiance becomes smaller. As the angle becomes perpendicular to the surface normal \vec{n} , no light is reaching the surface anymore. The relationship between the angle of incidence θ_i on a surface and irradiance is expressed by the cosine of θ_i . This relationship is known as Lambert's law (depicted in ??) and the equation for irradiance becomes:

$$E = \frac{\Phi \cos \theta_i}{A} \quad (2.2)$$

2.2.4 Radiance

Radiance L measures irradiance with respect to a solid angle ω as the solid angle goes to zero. A solid angle going to zero essentially means that the cone shrinks in size to the limit of 0. In that case the solid angle becomes a direction ω . In other words, radiance is defined as differential power $d\Phi$ per differential area dA at a point p for direction ω and can be written as [?]:

$$L(p, \omega) = \frac{dE_\omega(p)}{d\omega} = \frac{d\Phi \cos \theta}{d\omega dA} \quad (2.3)$$

Radiance L measures the light intensity per unit area on a surface. It measures the quantity of radiation that passes through or is emitted from a surface and falls within a given solid angle in a specified direction. Radiance is used to characterize diffuse emission. The unit of radiance is watts per steradian per square meter ($W \cdot sr^{-1} \cdot m^{-2}$).

A nice thing about radiance is that it remains constant along rays as long as the ray moves through a vacuum. This makes it very useful in rendering applications.

2.2.5 Bidirectional Reflection Distribution Function (BRDF)

2.3 $\sec_b sdf$

When a surface is lit by a light source, part of it will scatter back into the environment. The distribution at which light is scattered back is described by a bidirectional scattering distribution function (BSDF). It is named bidirectional, because the distribution is dependent on the incident and outgoing light direction. If the surface is opaque, which means there is no subsurface light transport or refraction, then the BSDF is described by surface reflection only. In that case the BSDF is called a bidirectional reflectance distribution function (BRDF). A BRDF function is denoted as $f_r(p, \omega_i, \omega_o)$ where p is the surface position, ω_i is the incident light direction and ω_o the outgoing light direction.

A BRDF can be expressed as a ratio between the outgoing radiance L_o in direction ω_o and an incident irradiance E_i from direction ω_i . Since we are dealing with a continuous flow of energy transport (there is no snapshot of light), the BRDF ratio is expressed as a differential, as shown in equation ??.

$$f_r(p, \omega_o, \omega_i) = \frac{dL_o(p, \omega_o)}{dE(p, \omega_i)} = \frac{dL_o(p, \omega_o)}{L_i(p, \omega_i) \cos \theta_i d\omega_i} \quad (2.4)$$

The rewritten form of irradiance in the equation above comes from the definition of irradiance as has been explained in section ??.

BRDF's have two important characteristics [?]:

- Reciprocity: the incoming and outgoing directions can be swapped and still have the same reflection behavior: $f_r(p, \omega_i, \omega_o) = f_r(p, \omega_o, \omega_i)$.
- Energy conservation: the total amount of energy that is reflected will always be less or equal than the amount of energy that is incident to the surface.

2.3.1 Bidirectional Curve Scattering Distribution Function (BCDSF)

As discussed in the previous section, a BRDF describes the ratio between outgoing surface radiance in direction ω_o , to incident surface irradiance from a differential solid angle ω_i [?]. Fibers are usually treated as one-dimensional curves. Curves have no area and therefore, scattering from fibers need to be described in terms of the length of a fiber.

Scattering from a fiber is thus represented by a bidirectional curve scattering distribution function (BCSDF) $S(\omega_i, \omega_o)$ that describes the ratio between the outgoing curve radiance L_o in direction ω_o to the incident curve irradiance E_i from a differential solid angle ω_i . It shares the same physical units as a BRDF, but the concept is a little different.

$$S(\omega_i, \omega_o) = \frac{dL_o(\omega_o)}{dE_i(\omega_i)} \quad (2.5)$$

Curve irradiance is proportional to incoming curve radiance:

$$dE_i(\omega_i) = DL_i(\omega_i) \cos \theta_i d\omega_i \quad (2.6)$$

Here D denotes the diameter of the fiber. So, curve irradiance is the amount of light falling on an infinitesimal small length of curve dl , times the diameter of the fiber: Ddl . Using the former equation, the scattering integral can be written as:

$$L_o(\omega_o) = D \int S(\omega_i, \omega_o) L_i(\omega_i) \cos \theta_i d\omega_i \quad (2.7)$$

2.3.2 Rendering Equation

The rendering equation is an integral and formally describes the amount of radiance leaving from a position p in direction ω_o . The radiance consists of the emitted radiance L_e plus the integral around the sphere to integrate all incoming light L_i from incident direction ω_i , reflecting in direction ω_o , where the BSDF is denoted by $f(p, \omega_o, \omega_i)$ indicating the contribution of incident light reflected in the outgoing direction at position p . The cosine factor is to take into account the angle with the surface normal (see section ??).

$$L_o(p, \omega_o) = L_e(p, \omega_o) + \int_{\Omega} f(p, \omega_o, \omega_i) L_i(p, \omega_i) \cos \theta_i d\omega_i \quad (2.8)$$

Solving this equation is the key task for a computer graphics program. Hair fibers do not emit radiance and therefore the emitted radiance can be ignored in the rendering equation. Taken into account that hair fibers are represented by curves (see section ??), the curve rendering equation that will be used in this thesis is as follows.

$$L_o(p, \omega_o) = D \cdot \int_{\Omega} S(p, \omega_o, \omega_i) L_i(p, \omega_i) \cos \theta_i d\omega_i \quad (2.9)$$

By removing the position parameter p , we obtain the same equation as equation ??.

2.4 Importance Sampling

Importance sampling is a variance reduction technique. The idea is to prefer samples that have more impact on the end result (i.e. the rendering) than samples that hardly have an influence. In other words, samples with a higher impact are more important to sample from, than samples that hardly have impact. By preferring samples of important directions we are converging faster to the correct end result, thereby decreasing rendering times significantly.

2.4.1 Monte-Carlo integration

Monte-Carlo integration is a numerical integration technique. In contrast to many other integration techniques, it can evaluate an integral using random samples, where performance is irrespective of the dimensionality of the samples.

Given an integral $\int_a^b f(x)dx$ for a function $f(x)$ over the range $[a, b]$. If it is possible to find a function $F(x) = \int f(x)dx$, then we could solve the integral analytically and there would be no need for numerical integration.

In the context of physically based rendering, the integral is often too complicated to solve analytically. The Monte-Carlo integration is a relatively simple way to be able to find the expected value of the integral $E[F(x)]$. Monte-Carlo integration is doing that by taking N random samples where each sample is weighted by the probability of being sampled $p(X)$. The larger the number of samples, the more accurate the result will be, eventually converging to the correct result for the integral. The following equation describes the computation of the integral F_N with the Monte-Carlo estimator [?].

$$F_N = \frac{1}{N} \sum_{i=1}^N \frac{f(X_i)}{p(X_i)} \quad (2.10)$$

N represents the number of samples taken, X_i represents the i -th random sample and $p(X_i)$ is the probability density for sample X_i . If the probability density function $p(x)$ is uniform then $p(x)$ becomes a constant $\frac{1}{c}$ and can be moved to the front of the equation, resolving to the following Monte-Carlo estimator.

$$F_N = \frac{c}{N} \sum_{i=1}^N f(X_i) \quad (2.11)$$

Integrating over a range $[a, b]$ resolves to a constant $c = b - a$ and thus it boils down to finding the average value for $f(x)$ multiplied by the range, giving the area under the surface or when integrating around a hemisphere, the solution to the rendering equation.

2.4.2 Importance sampling

It becomes interesting when the probability density function $p(x)$ is not constant for all samples. By giving certain samples a higher probability and other (less important) samples a lower probability, we can converge faster to the correct result by using less samples.

The idea behind importance sampling is to draw samples from the probability distribution function (PDF). There are different techniques to draw samples from a PDF as explained in the book by Pharr et al. [?]. Possible techniques are inversion sampling, in which the normalized cumulative distribution function (CDF) is computed, then inverted and used with a uniform random variable $\xi \in (0, 1]$ to obtain a sample X . Other techniques are rejection sampling.

It is easy to see that the better the fit of the PDF with the function $f(x)$, the better the drawn samples and the faster we converge to the correct result. If $f(x) = p(x)$ we are talking about perfect importance sampling, the most ideal case. However, it is not always possible to invert $f(x)$. If we can find a PDF that is still a close fit to $f(x)$ but is possible to invert, then we still have a better sampling technique compared to uniform sampling.

Other techniques are possible as well, such as rejection sampling, which is described in detail in Pharr et al. [?].

2.4.3 Multiple importance sampling

Sometimes we end up with a function that can be written as the product of multiple functions $f(x)g(x)$, each of which has an easy to sample PDF $p_f(x)$, belonging to $f(x)$ and $p_g(x)$ for $g(x)$.

If two sampling distributions $p_f(x)$ and $p_g(x)$ are used to estimate the value of $\int f(x)g(x)dx$, you cannot just average the results by taking N samples from $p_f(x)$ and M samples from $p_g(x)$ and average the result, because variance is additive. The way to do it is by using the new Monte-Carlo estimator for multiple importance sampling (MIS) stated in [?]:

$$\frac{1}{n_f} \sum_{i=1}^{n_f} \frac{f(X_i)g(X_i)w_f(X_i)}{p_f(X_i)} + \frac{1}{n_g} \sum_{j=1}^{n_g} \frac{f(Y_j)g(Y_j)w_g(Y_j)}{p_g(Y_j)} \quad (2.12)$$

where w_f and w_g are weighting functions.

$$w_s(x) = \frac{n_s p_s(x)}{\sum_i n_i p_i(x)} \quad (2.13)$$

An example of multiple importance sampling is when rendering complex scenes with multiple objects and light sources. When light sources are very small, the chance of being sampled (as observed from a point on a surface) is likewise very small. It is easy to imagine that lots of samples are required when evaluating the BSDF for a material to remove the noise. A better approach would be to sample from the light source. If the light source is very large, for example a skydome around the scene acting as a light source, it is more efficient to sample the BSDF of the material. Multiple importance sampling takes this effect into account. By mixing samples from the BSDF of the material and from the distribution of the light source, a selection of important samples can be used to reduce noise.

Chapter 3

Related Work

3.1 Hair Rendering Overview

The explicit representation of hair provides the most physically accurate results. In this thesis, explicit representations are used. As will be explained, the Marschner model is an accurate single fiber scattering model. To increase the realistic appearance of hair volumes, more fibers should be taken into account. This leads to the concept of multiple fiber scattering. Multiple fiber scattering is analogous to subsurface scattering in solid materials and therefore necessary to increase the level of realism. Ward et al. [?] provides a good overview on the different models in hair rendering in general.

In this section, related work regarding the development of hair models is explained, followed by an in-depth treatment of the most relevant work related to the research in this thesis.

3.1.1 Single fiber scattering

Kajiya and Kay [?] provided one of the first widely used hair scattering model that was designed to render fur. Fur usually refers to a dense coat of fine and soft hair. Their idea was that fine hair geometry should not be represented with complex geometry which leads to aliasing effects, but with texels. A texel is a three-dimensional texture map intended to represent a highly complex collection of surfaces contained within a defined volume [?].

The rendering of the hair model was based on a diffuse and specular component. The diffuse component is obtained by simply integrating a Lambert surface along one half of the cylinder facing the light source. The specular component is based on the Phong model in which light is reflected at a mirror angle along the tangent. Since the normals on the cylinder point in directions perpendicular to the tangent, the reflected light forms a cone whose angle at the apex is equal to the angle of incidence [?]. The Kajiya and Kay model is clearly an ad-hoc model where hair fibers are treated as



Figure 3.1: Comparison of Kajiya and Kay model with the Marschner model. The left image represents Kajiya and Kay’s model. The center image is the Marschner model and the right image represents the original real-world photographed reference. Picture taken from Marschner et al. [?].

opaque cylinders, limiting the realism of the model (see figure ??).

The Kajiya and Kay model lacks directionality, meaning that hairs are fully lit independent of the light direction or eye position. Goldman [?] was interested in reflection and transmission and increased directionality by introducing two new attenuation factors. These attenuation factors can be used to tune the relative transmissivity and reflectivity of a hair. It does this by computing a value f_{DIR} that is multiplied with the contribution of the Kajiya and Kay model.

Marschner et al. [?] improved the Kajiya and Kay model substantially by proposing the most physically based hair scattering model at that time. Their model makes two improvements to Kajiya and Kay’s model: it predicts the azimuthal variation in scattered light based on the ray optics of a cylinder, and it accounts for the longitudinal separation of the highlight into surface-reflection, transmission, and internal-reflection components that emerge at different angles [?]. The Marschner model is described in detail in section ??.

3.1.2 Multiple fiber scattering

Ward et al. [?] states that for light-colored hair, recent work has shown that shadowing and attenuation alone are insufficient to produce the correct appearance. For fully realistic results multiple scattering must be accounted for.

Multiple fiber scattering extends single fiber scattering to take into account the scattering effects taking place between multiple hair strands. A single fiber scattering model can be used as the basis for multiple fiber scattering. For example, if the Marschner model is the basis for single fiber

scattering, then in combination with path tracing the result is obtained for multiple fiber scattering. Path tracing is a brute-force approach that gives realistic results. However, given the geometric complexity of a hair model, it is a very time and memory intensive process. More efficient models for multiple fiber scattering models should be devised.

Photon mapping methods can reduce per-frame rendering times from days, required for path tracing methods, to hours [?]. Photon mapping is a practical approach for computing global illumination in complex environments. Photon mapping is introduced by Henrik Wann Jensen [?] and it is, in general, a two-step approach. The first pass in the method is constructing the photon map by emitting photons from the light sources in the model and storing these in the photon map as they hit surfaces [?]. The collection of photons distributed in the photon map, can be seen as a rough representation for the amount of light in the scene. The second step is that this information from the photon map is used to render the scene.

For multiple scattering in hair models, Moon and Marschner [?] developed a model using a photon mapping approach. In the first pass, particles are traced from light sources into the hair volume and followed through multiple scattering events, and their positions and directions are stored into a 5D hierarchical data structure to record the flow of particles through space [?]. In the second pass a density estimate is performed simultaneously in position and direction to estimate the radiance arriving at a hair from a particular direction. This photon mapping approach leads to similar results compared to path tracing, but the biggest drawback is that it requires high resolution photon maps (memory usage) and the method is not interactive.

The Dual Scattering model by Zinke et al. [?] is a direct extension on top of Marschner and doesn't have the drawbacks of photon mapping approaches. The dual scattering approximation splits up the compensation in a global and a local scattering component. The global scattering component computes the amount of irradiance reaching a certain position in the hair model. The local scattering component uses the computed irradiance to compute the scattering effects in the direct neighbourhood of the position to be rendered. More in-depth theory is provided in section ??.

3.2 Marschner model

The biggest drawback of Kajiya and Kay [?] is that their scattering model is not based on physical measurements. Instead, it uses a simple ad-hoc term with a constant linear highlight. Marschner et al. [?] proposed a physically based scattering model for single fiber scattering. The single fiber scattering model, or Marschner model, can be considered the status quo in modern hair rendering technology. Most physics based hair rendering model in some way

or another are derived from the Marschner model. Likewise, the Marschner model forms the basis for the dual scattering approximation (by Zinke et al. [?]). In this section the Marschner model will be discussed in detail.

3.2.1 Observations

In order to come up with a physically based scattering model for single hair fibers, Marschner et al. devised a setup to capture the response of light scattered against a hair fiber.

They illuminated individual hairs with a narrow beam and measured the scattered light in various directions, using a setup based on a four-axis goniometer that positioned a light source and a CCD camera at arbitrary directions from the sample [?]. Their measurements show a couple of distinct characteristics for human hair.

Longitudinal variation (Primary and secondary highlight)

Stamm et al. [?] and Bustard and Smith [?] measured the scattering behaviour in the incidence plane. This is the 2D slice for which the hair fiber is coplanar with the source and detector. Marschner et al. did the same measurements and placed a light source under a direction of positive and negative 45 degrees with the hair fiber ($\theta_i = +/ - 45^\circ$). See figure ?? for the measured responses.

Observations show that synthetic hair has a highlight that appears exactly at the specular direction ($\theta_r = -\theta_i$), but for real hair fibers the specular highlight is shifted toward the root of the hair fiber. This is due to the tilted cuticle scales that have a surface orientation that is shifted by approximately 6 to 8 degrees from the ideal cylinder surface normal. This highlight is the result of scattering against the surface of the hair fiber. It is called the primary specular highlight and appears as having the color of the light source.

Another observation from figure ?? is that blond hair has a coloured secondary peak. This can be seen by the red response that is slightly higher compared to the blue and red response. The secondary peak can be explained by internal scattering. Consider the hair fiber as a cylinder; when light enters the cylinder, the light will first refract, then propagates through the core of the fiber and bounces back against the other end of the fiber. When leaving the fiber, the light refracts again. This reflection and refraction behaviour causes the deviation from the primary specular highlight. During its travel, light propagated through the center of the hair fiber, where the pigments are. These pigments absorb part of the energy in different wavelengths. This is how the secondary peak obtains its color.

Furthermore, Marschner et al. [?] noticed that as the scattering angle increases, the secondary highlight fades out, while the primary highlight maintains more constant amplitude. Both peaks maintain approximately

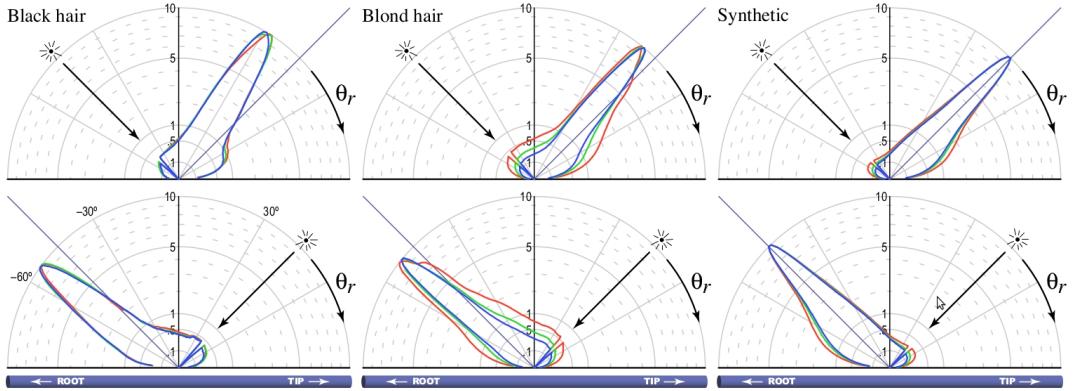


Figure 3.2: The scattering behaviour in the incidence plane (longitudinal scattering). For human hair the specular reflection is shifted towards the root by a few degrees. For synthetic hair no shifted highlight is visible, because synthetic hair doesn't consist of tilted cuticle scales.

constant width and at glancing angles the primary highlight becomes a sharp peak that is prominent very close to the specular direction.

Azimuthal variation

Azimuthal variation is measured by placing the light and detector in the plane that is perpendicular to the hair fiber: the normal plane. In figure ?? plots are displayed that show the azimuthal scattering response with varying ϕ_i and ϕ_r . The setup is that the light source is fixed (targeted from right to left) and the hair fiber is rotated.

In this normal plane setup, it is clearly visible that there are two bright out-of-plane peaks. Out of plane, because the response is captured at $\theta_r = 10$ degrees, due to the tilted cuticle scales discussed before. These peaks are called glints.

Glints change considerably in brightness and position as a function of θ_i , meaning that the hair is not rotationally symmetric [?]. This is because the hair is eccentric and has an elliptical cross section. More generally, the evolution of the peaks as the fiber rotates appears similar to the internal reflection from a transparent elliptical cylinder [?].

Figure ?? also shows a strong transmission component. This is light that passes through the hair fiber and leaving on the other side of the fiber as seen from the light source. This means that the Marschner hair model takes three scattering modes into account. R stands for reflection and T for transmission:

- R: Specular reflection that is deviated slightly from the perfect specular

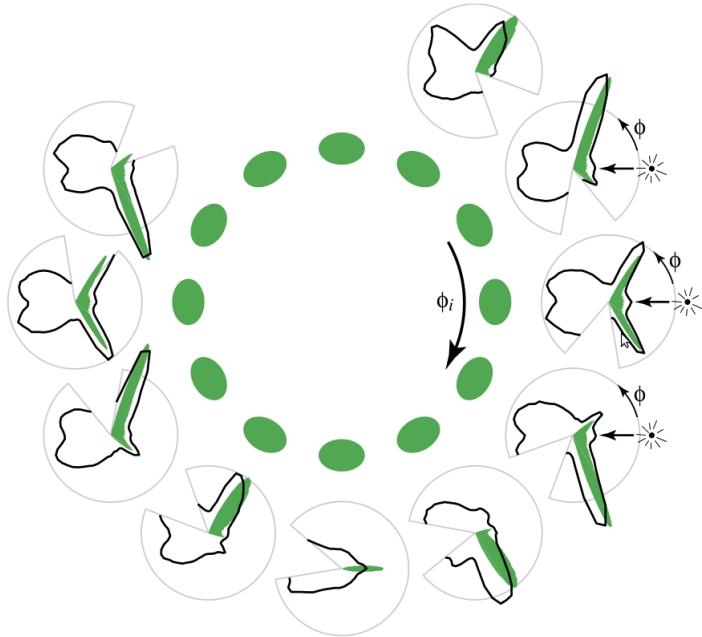


Figure 3.3: Measurements of azimuthal scattering (scattering in the normal plane). The light source is shining from the right, while the hair fiber is rotated. The green ellipse is the cross section of the fiber for different measurements with increasing rotations of the fiber. Figure taken from Marschner et al. [?].

direction, due to the cuticle scales.

- TT: Transmission component that enters the fiber, propagates through it and leaves the fiber at the other end.
- TRT: Internal reflection, that enters the fiber, scatters back and leaves again at the same side as where it entered.

More scattering modes are possible as well, such as TRRT or TRRRT. These events are ignored, because their contributions become negligible small due to loss of energy after that many reflection events. Instead, Marschner et al. [?] suggests to add a little diffuse color for better looking results. In figure ?? an image is shown with the three scattering components clearly visible. D'Eon et al. [?] generalized the Marschner model to take into account all possible scattering events, so that no energy is lost. This is mainly important for highly transmissive hair fibers, such as white hair for polar bears. For human hair-rendering this might not directly be an issue.



Figure 3.4: An image from Gray et al. [?] showing a female with brown hair. The three scattering modes: R, TT and TRT contributions are visible in the hair. Glints give hair it's characteristic texture.

3.2.2 Model

There are properties that have been used in previous work on scattering from fibers (Marcuse [?], Adler et al. [?], Mount et al. [?]):

- A ray that enters a dielectric cylinder at a particular angle to the axis will always exit at the same angle, regardless of the sequence of reflections and refractions it undergoes.
- The dependence of the scattered distribution can be analysed by examining only the projection into a plane perpendicular to the hair.

Based on the observations described in the previous subsection and the theory here, Marschner et al. came up with a model that is divided into a longitudinal scattering function M_p and an azimuthal scattering function N_p , where p is the scattering mode with $p \in \{R, TT, TRT\}$. The resulting scattering model can then be written as:

$$S(\omega_i, \omega_r) = \frac{\sum_{p \in \{R, TT, TRT\}} M_p(\theta_i, \theta_r) N_p(\eta'(\theta_d); \phi_i, \phi_r)}{\cos^2 \theta_d} \quad (3.1)$$

S depends on incoming and outgoing angles ω_i and ω_r , but can be reduced to a 2D scattering function, since the scattering model makes use of derived angles θ_d , θ_h and the relative difference in azimuthal angle ϕ (see section ??).

Longitudinal Scattering Function

Using the first property of scattering from fibers, it can be derived that the angle of incidence is the same as the angle of reflection ($\theta_i = -\theta_r$). There are two deviations that cause the reflection to not be exactly in the specular direction (see section ??). First, the surface of a hair fiber is rough, causing a spread of reflected light instead of a mirror like specular reflection. Secondly, the cuticle scales of the hair fiber are tilted, causing a 6 to 8 degrees deviation from the perfect specular direction.

Marschner et al. [?] approximates the longitudinal scattering function M by a unit-integral zero-mean gaussian function $g(\beta, x)$, where β equals the standard deviation and α equals the longitudinal shift. D'Eon et al. [?] observed that the Marschner model is not entirely energy conserving for several reasons. One of the reasons is that the use of the half angle θ_h doubles the reflected energy on average. Additionally the gaussian is normalized with respect to $\theta_h \in (-\infty, \infty)$, but is evaluated for $\theta_h \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, thereby losing a considerable amount of energy that is present outside of the used range. The doubling of energy is easily accounted for by doubling the half angle in the Marschner longitudinal scattering function. The resulting longitudinal functions M that are used for the Marschner model are thus as follows:

$$M_R(\theta_h) = g(\beta_R, 2\theta_h - \alpha_R) \quad (3.2)$$

$$M_{TT}(\theta_h) = g(\beta_{TT}, 2\theta_h - \alpha_{TT}) \quad (3.3)$$

$$M_{TRT}(\theta_h) = g(\beta_{TRT}, 2\theta_h - \alpha_{TRT}) \quad (3.4)$$

By varying the standard deviation β , the spread of the longitudinal scattering lobe can be adjusted to simulate the rough surface of hair fibers. The longitudinal shift α deviates the peak from the specular direction by shifting the gaussian distribution. Some typical values for α and β are presented in table ??.

Azimuthal Scattering Function

For the azimuthal scattering component different scattering paths inside the circular cross section of the hair need to be considered. The Marschner model takes into account 3 scattering modes, the R, TT and TRT mode. Figure ?? shows a graphical representation of a circular cross section of a hair fiber.

Parameter	Description	Typical value
α_R	Longitudinal shift for R lobe	-10° to -5°
α_{TT}	Longitudinal shift for TT lobe	$-\alpha_R/2$
α_{TRT}	Longitudinal shift for TRTlobe	$-3\alpha_R/2$
β_R	Longitudinal spread for R lobe	5° to 10°
β_{TT}	Longitudinal spread for TT lobe	$\beta_R/2$
β_{TRT}	Longitudinal spread for TRT lobe	$2\beta_R$

Table 3.1: Typical values for parameters for the longitudinal scattering function. α is the longitudinal shift to represent the shift of the lobe due to tilted cuticle scales. β is the longitudinal width, representing the spread of the energy due to rough surface of hair fibers. Taken from Marschner et al. [?]

Scattering from a dielectric circle is well studied. Consider a parallel beam (or ray) incident to a circular cross section at an offset h from the center. A parallel beam incident to the center of a circle corresponds to the root value $h = 0$. A value of $h = 1$ or $h = -1$ corresponds to beams glancing the edges of the circle. Using the offset h the angle of incidence γ_i can be computed using the following equation:

$$\begin{aligned}\sin \gamma_i &= h \Leftrightarrow \gamma_i = \arcsin h \\ \eta \sin \gamma_t &= h \Leftrightarrow \gamma_t = (\arcsin h)/\eta\end{aligned}$$

A ray that is incident to the unit circle can be traced as it refracts and reflects through the circle. The exit angle is not only dependent on the incidence angle (or offset from the center), but also on the number of scattering events. The number of scattering events p equals 1 for the reflection case, 2 for transmission (TT) and 3 for internal reflection (TRT). The exit angle can then be calculated as follows:

$$\phi(p, h) = 2p\gamma_t - 2\gamma_i + p\pi \quad (3.5)$$

Finding roots

Consider rendering a scene using ray tracing and that a particular camera ray is incident to a hair fiber. Figure ?? represents this situation. To find the contribution of radiance towards the camera (i.e. the outgoing direction ω_o), all the paths that contribute to scattering in that direction should be taken into account. These incident light directions ω_i can be integrated around the sphere by uniform sampling (or importance sampling). Given two directions the contributed radiance can be found by using the difference angle $\phi = \phi_o - \phi_i$.

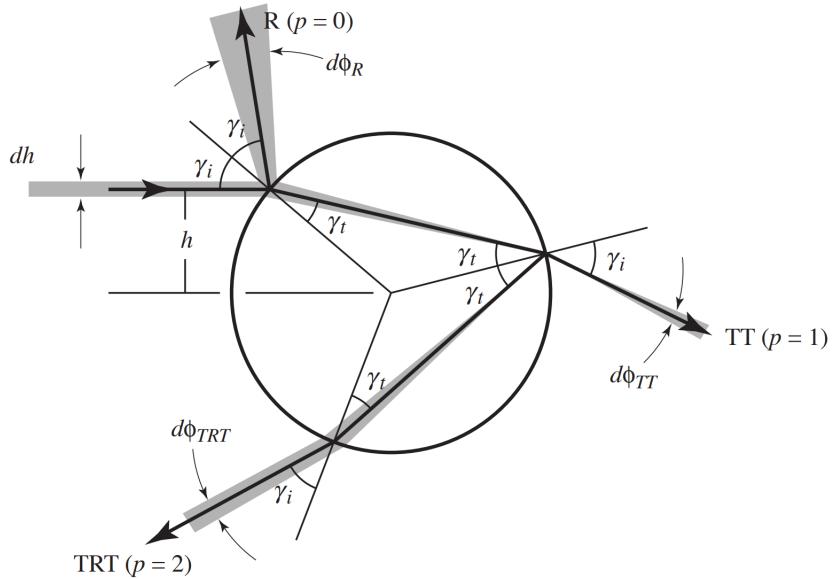


Figure 3.5: Cross section of a hair fiber showing the different scattering modes. Taken from Marschner et al. [?]

Looking again at equation ??, the exit angle $\phi(p, h) = \phi$ is known. To be able to evaluate the Marschner model, the roots h must be found. This is done by solving the roots by solving for the roots of the function $\phi(p, h) - \phi = 0$, which gives γ_i which trivially gives h .

To find the contribution of radiance from the incident direction ω_i and ω_o the roots have to be found for all scattering modes in order to evaluate the rendering equation in equation ???. Finding the roots requires solving a cubic equation, which is explained deeper in Marschner et al. [?]. The R and TT cases have exactly one root and TRT has either one or three roots.

Attenuation

The light that propagates through a hair fiber is attenuated. Attenuation occurs because of reflections, refractions and absorption. The attenuation for reflections and refractions are computed using the Fresnel equation. The Fresnel equation takes a parallel and perpendicular index of refraction (η' and η'' respectively). Absorption takes place due to the pigments in the core, giving hair its color. The amount of absorption depends on the length of each internal path segment l that is computed by applying the law of cosines. According to d'Eon et al. [?], the formula given in Marschner et al. [?] is wrong. The equation to find the path length of a path segment l should be as followed:

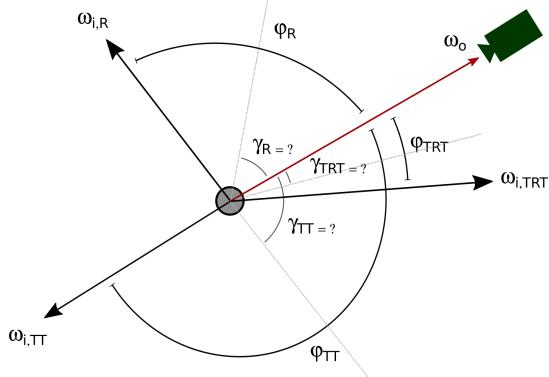


Figure 3.6: Visualization of the root finding stratgegy. The relative difference angle ϕ_p between rays ω_o and $\omega_{i,p}$ is known by computing the difference of the azimuthal spherical coordinates. The Marschner model can be evaluated when the roots are known. The roots h can be found by solving the roots of equation $2p\gamma_t - 2\gamma_i + p\pi - \phi_p = 0$. Knowing γ_i will trivially give h .

$$l = \frac{2r \cos \gamma_t}{\cos \theta_t} \text{ with } \theta_t = -\operatorname{sgn}(\theta_i) \arccos((\eta''/\eta') \cos \theta_i) \quad (3.6)$$

Here l denotes the length of a single internal path segment for the TT case. For TRT, the length doubles. r is the radius of the hair fiber and is a setting provided by the user. Absorption is denoted by $T(\sigma_a, h)$ where σ_a is the absorption coefficient. It describes the amount of absorption per unit length through the hair fiber. It follows trivially that the absorption is computed as $T(\sigma_a, h) = \exp(l \cdot \sigma_a)$. The normal plane scattering function N_p can now be described and is shown in equation ??, where p corresponds to the scattering modes R, TT, TRT with $p = 0, p = 1, p = 2$ respectively.

$$N_p(\phi) = \sum_r A(p, h(p, r, \phi)) \cdot \left| 2 \frac{d\phi}{dh}(p, h(p, r, \phi)) \right|^{-1} \quad (3.7)$$

with $h(p, r, \phi)$ corresponding to the roots r for scattering mode p and difference angle ϕ . The attenuation $A(p, h)$ is described as:

$$A(0, h) = \operatorname{Fresnel}(\eta, \gamma_i) \quad (3.8)$$

$$A(p, h) = (1 - \operatorname{Fresnel}(\eta, \gamma_i))^2 \cdot \operatorname{Fresnel}\left(\frac{1}{\eta}, \gamma_t\right)^{p-1} \cdot T(\sigma_a, h) \quad (3.9)$$

Glints

Because the theory is based on smooth surfaces, the normal plane scattering function N_p for the TRT scenario produces singularities with infinite intensity [?]. These singularities are called caustics. Caustics are bright spots that originate when light bundles together after a sequence of reflections and refractions. Caustics can be observed found under water or when looking through a wine glass for example. These infinite singularities are removed by replacing it with a smooth gaussian lobe centered at the location of the caustic. This changes the normal plane scattering function for the TRT case. Marschner et al.[?] provides more detail into how the scattering function changes.

Taking into account eccentricity

Changing refractive index has effects that are qualitatively similar to changing eccentricity [?]. In order to simulate eccentricity, the refractive index η^* is computed based on the value of eccentricity a . The equation can be found below. More detail is in Marschner et al's paper [?].

$$\eta_1^* = 2(\eta - 1)a^2 - \eta + 2 \quad (3.10)$$

$$\eta_2^* = 2(\eta - 1)a^{-2} - \eta + 2 \quad (3.11)$$

$$\eta^*(\phi_h) = \frac{1}{2}((\eta_1^* + \eta_2^*) + \cos(2\phi_h)(\eta_1^* - \eta_2^*)) \quad (3.12)$$

The Marschner scattering model

From the description above we can now describe the complete Marschner scattering model $S(\omega_i, \omega_r)$.

$$\begin{aligned} S(\phi_i, \theta_i; \phi_r, \theta_r) &= M_R(\theta_h)N_R(\phi)/\cos^2\theta_d \\ &+ M_{TT}(\theta_h)N_{TT}(\phi)/\cos^2\theta_d \\ &+ M_{TRT}(\theta_h)N_{TRT}(\phi)/\cos^2\theta_d \end{aligned}$$

3.2.3 Evaluation

The advantage of the Marschner hair model is that it is based on physical measurements. Compared to previous work it comes much closer to physical reality. Finding the roots for a given direction is tedious. d'Eon et al. [?] proposed a faster alternative where root-solving is not needed anymore. Also some flaws in the Marschner model, concerning energy conservings are removed.

The Marschner hair model focuses on single fiber scattering and to render realistic hair volumes, multiple fiber scattering needs to be taken into account. Path tracing or photon mapping are possible approaches, but take a lot of time or require an extensive amount of memory. The dual scattering method by Zinke et al. [?] proposes a faster way where extensive path tracing is not needed anymore, drastically speeding up render process. This model is the topic of the next section.

3.3 Dualscattering Approximation

Multiple scattering is a key factor in the realistic rendering of hair, especially for dense light colored hair volumes (such as blond hair). The relative amount of absorption by light colored hair types is fairly low, so this means that light undergoes many scattering interactions inside the hair volume before the contribution is reduced to an amount that can be negligible. The high geometric complexity of hair models coupled with the complexity of the light interaction in hair volumes makes computing the multiple scattering effect difficult [?]. Difficult to an extent that path-tracing is not a feasible solution, because it requires too much time and too many samples to reduce the noise and come to an acceptable rendering.

The dual scattering approximation, proposed by Zinke et al. [?] splits the multiple scattering computation in two components: global multiple scattering and local multiple scattering. The global multiple scattering component aims to compute the light traveling through the hair volume and reaching the neighborhood of the point of interest, while local multiple scattering accounts for the scattering events within this neighborhood [?].

Starting from the general rendering equation (see equation ??) and taking the diameter D as being 1, we obtain the following equation.

$$L_o(x, \omega_o) = \int_{\Omega} L_i(x, \omega_i) S(\omega_i, \omega_o) \cos \theta_i d\omega_i \quad (3.13)$$

This can be interpreted as the radiance in outgoing direction ω_o at a position x , equals the integral around the sphere (Ω) of incident directions ω_i multiplied by the BSDF factor. The integration is a recursive process, and to find the incident radiance $L(x, \omega_i)$, another integration is needed from another point x' to find $L_o(x', -\omega_i)$. The dualscattering approximation simplifies this process by defining the incident radiance as

$$L_i(x, \omega_i) = \int_{\Omega} L_d(\omega_d) \Psi(x, \omega_d, \omega_i) d\omega_d \quad (3.14)$$

where L_d is the incident radiance from outside the hair volume from direction ω_d (assuming distant illumination), and $\Psi(x, \omega_d, \omega_i)$ is the multiple scattering function denoting the fraction of light entering the hair volume

from direction ω_d that is scattered inside the hair volume and finally arriving at point x from direction ω_i [?]. This works similar as to how subsurface scattering is defined.

The main concept behind the dual scattering method is to approximate the multiple scattering function as a combination of two components. The global multiple scattering function Ψ^G is used to compute the irradiance arriving at the neighborhood of point x inside the hair volume, and the local multiple scattering function Ψ^L approximates the multiple scattering of this irradiance within the local neighborhood of x [?]. This results in the following equation, central to the dual scattering approximation model.

$$\Psi(x, \omega_d, \omega_i) = \Psi^G(x, \omega_d, \omega_i)(1 + \Psi^L(x, \omega_d, \omega_i)) \quad (3.15)$$

The global multiple scattering is thus responsible for gathering the irradiance from outside the hair volume and taking into account the scattering effects in the hair volume, to deliver an approximate amount of irradiance arriving at point x . It is important to see that the global multiple scattering removes the need to do extensive path tracing through the hair volume.

The local multiple scattering is more in line with the Marschner model: given the irradiance at point x , calculate the scattering effect of the local neighbourhood. For a neighborhood consisting of a single fiber, it boils down to evaluation of the Marschner model.

3.3.1 Global Multiple Scattering Function Ψ^G

The global multiple scattering function Ψ^G gathers the irradiance from outside the hair volume, approximating the amount of irradiance arriving at a point x inside the hair volume. The global multiple scattering is defined as follows by Zinke et al. [?]:

$$\Psi^G(x, \omega_d, \omega_i) \approx T_f(x, \omega_d) S_f(x, \omega_d, \omega_i) \quad (3.16)$$

The global multiple scattering Ψ^G is the multiplication of forward transmittance T_f and the average forward scattering spread S_f . The transmittance accounts for the amount of hairs in between the light source and the point to be shaded. The average forward scattering spread takes into account the loss of energy due to spreading of the light after each scattering event.

Forward Scattering Transmittance

In this thesis the forward scattering transmittance is computed by using a combination of ray shooting together with a voxel grid. Before rendering, a voxel grid is generated based on the hair model. Each voxel cell contains the density of the hairs in question. This is computed by traversing each

hair fiber and storing the distance of the hair fiber segment in the voxel cell. In this way, cells that contain a lot of hair fibers, do have a higher density value.

When finding the irradiance at a position x in direction ω_i , a ray is shot in the direction of the light source. This ray is linearly ray-marched and for each position along the ray, the density value is looked up in the voxel cell. In this way, the global density value is computed for a specific direction ω_i . The density factor can then be translated to the number of hair strands, by dividing the density by the volume of a voxel cell. This results in n , the number of hair strands in between the shading point x and the light from direction ω_i .

The forward scattering transmittance for the dual scattering method [?] is defined as follows:

$$T_f(x, \omega_d) = d_f(x, \omega_d) \prod_{k=1}^n a_f(\theta_d^k) \quad (3.17)$$

where d_f is the forward scattering density factor, which is a constant between $[0, 1]$ and usually set to 0.7. This factor is a tweaking variable that can be changed dependent on the denseness of a hair volume. See figure ??.

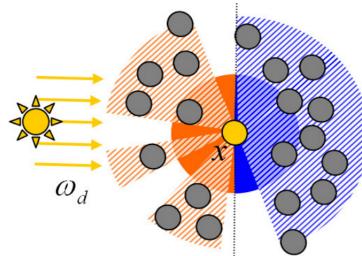


Figure 3.7: This picture shows the cross section of a hair cluster. The factor d_f is a tweaking variable to account for the portion of multiple scattered radiance reaching the point x . The orange region shows the forward scattering contribution. It can be seen that the light could have direct vision with point x in which case there is no multiple scattered radiance. d_f accounts for this fact. If d_f is 1, then it is assumed that point x is fully covered by hair strands as seen from the light direction. The same applies for the backward scattering attenuation factor d_b (blue region). Usually both d_f and d_b are constant for the full model and set to 0.7.

$a_f(\theta_d^k)$ is the average forward scattering attenuation at the k -th scattering event, which is computed in a precomputation step by integrating over all possible directions in the hemisphere and averaging the amount of forward scattered radiance. See the dual scattering method for more details [?] or

the efficient implementation of the dual scattering model in RenderMan by Sadeghi and Tamstorf [?] for a more hands on approach.

In this work, a simplification is used by usng a constant orientation for all k scattering events, under the assumption that we are dealing with human hair fibers that are locally similar in orientation. This means equation ?? can be simplified to the following equation.

$$T_f(x, \omega_d) = d_f(x, \omega_d) a_f(\theta_d)^n \quad (3.18)$$

Forward Scattering Spread

The spread function approximates the final angular distribution of front scattering light to find the probability of radiance coming to the point x from direction ω_i [?]. Because of the wide azimuthal scattering property of hair fibers, front scattered radiance quickly becomes isotropic in the azimuthal direction after only a few scattering events. For longitudinal scattering front scattered light is still anisotropic. The spread function is therefore split up into an isotropic azimuthal component ($1/\pi$) and an anisotropic longitudinal component described by a gaussian function g .

$$S_f(x, \omega_d, \omega_i) = \frac{s_f(\phi_d, \phi_i)}{\cos \theta_d} g(\theta_d + \theta_i, \sigma_f^2(x, \omega_d)) \quad (3.19)$$

where s_f is $1/\pi$ for forward scattering directions and zero for backward scattering. $\sigma_f^2(x, \omega_d)$ is the total variance of forward scattering in the longitudinal direction. This is defined by the Marschner model as the squared standard deviation β^2 and is dependent on the longitudinal angle. Simplifying again by taking a constant θ_d gives us a single multiplication of number of hair strands multiplied by the longitudinal variance.

$$\sigma_f^2(x, \omega_d) = \sum_{k=1}^n \beta_f^2(\theta_d^k) \approx n \cdot \beta_f^2(\theta_d) \quad (3.20)$$

3.3.2 Local Multiple Scattering

The local multiple scattering function Ψ^L accounts for the multiple scattering events within the neighborhood of the point x . Since light paths that go through only forward scattering are included in the global multiple scattering function, light paths of the local multiple scattering function must include at least one backward scattering.

Because of backward scattering, the local multiple scattering is mostly smooth with subtle changes over the hair volume [?].

The dual scattering model the local multiple scattering and the BCSDF of the hair fibers, approximating the result with a density factor d_b and backscattering function.

$$\Psi(x, \omega_d, \omega_i) f_s(\omega_i, \omega_o) \approx d_b(x, \omega_d) f_{\text{back}}(\omega_i, \omega_o) \quad (3.21)$$

The density factor d_b is a constant in the range $[0, 1]$, usually set to 0.7 approximating the density of the hair volume. The denser the hair, the stronger the backscattering event contributes to the result.

The backscattering is approximated with a material function f_{back}

$$f_{\text{back}}(\omega_i, \omega_o) = \frac{2}{\cos \theta_d} A_b(\theta_d) S_b(\omega_i, \omega_o) \quad (3.22)$$

Average Backscattering Attenuation

The average backscattering attenuation takes into account the attenuation in the local neighborhood of a point x . Since the global multiple scattering component already takes into account the forward scattering (meaning no backscatter events), we only need to deal with at least one backscatter event. Light coming from a light source, passing through the fiber at point x , needs an odd number of backscattering events for light to travel back to point x . The even number of backscattering events are ignored, since they are not directed to point x . Also, the dual scattering method ignores the contribution for more than three backscattering events, because their contribution becomes negligible low to the rendering result.

If we denote $a_b(\theta_d)$ as the average backscattering attenuation and $a_f(\theta_d)$ as the average forward scattering attenuation (described in the previous section), then the average backscattering for a single backscattering event A_1 can be computed as the attenuation of a single backscatter event, multiplied by the sum of the infinite series of forward scattering attenuations. The same holds for the average backscattering attenuation for three backscattering events A_3 , but now the possible paths are not a single sum anymore, but multiple sums to obtain the permutation of possible paths, dependent on where the backscattering event occurs in the sequence of forward scattering attenuations.

$$A_1(\theta_d) = a_b \sum_{i=1}^{\infty} a_f^{2i} = \frac{a_b a_f^2}{1 - a_f^2} \quad (3.23)$$

$$A_3(\theta_d) = a_b^3 \sum_{i=1}^{\infty} \sum_{j=0}^{i-1} \sum_{k=j+1}^{\infty} a_f^{2(i-j-1+k)} = \frac{a_b^3 a_f^2}{(1 - a_f^2)^3} \quad (3.24)$$

$$A_b(\theta_d) = A_1(\theta_d) + A_3(\theta_d) \quad (3.25)$$

where A_b is the average backscattering attenuation for 1 and 3 backscattering events. Figure ?? provides a graphics representation of the backscattering events and how they are directed with respect to position x .

(picture here)

Average Backscattering Spread

The average backscattering spread takes into account the loss of intensity, due to spreading of the light after backscattering events occur. This is similar to forward scattering events. The dual scattering method [?] presents the average backscattering spread $S_b(\omega_i, \omega_o)$ as

$$S_b(\omega_i, \omega_o) = \frac{s_b(\phi_i, \phi_o)}{\cos \theta} g(\theta_o + \theta_i - \Delta_b(\theta_d), \sigma_b^2(\theta_d)) \quad (3.26)$$

where s_b equals $1/\pi$ for backward scattering directions and zero for forward scattering, $\Delta_b(\theta_d)$ is the average longitudinal shift caused by the scattering events.

The spread is modeled as a single gaussian function. It is important to notice that Δ_b represents the average mean and σ^2 represents the average variance, thereby taking into account the multitude of possible forward and backward scattering events.

The average longitudinal shift Δ_b and average longitudinal standard deviation σ_b for backscattering can be precomputed by averaging over all directions. See the dual scattering method [?] for more details and [?] for a more hands on approach.

Chapter 4

Approach

The central problem in this thesis, and for rendering in general, is to improve rendering speed. Either by requiring less samples per pixel for the same quality image, or by increasing the quality of the image with the same amount of pixels. In both cases, it boils down to the same problem and this problem is usually solved by careful sample placement. Increasing the rendering speed is relevant for several applications, such as for the animation movie industry that wants to generate physically accurate movies, and for rendering real-time hair simulations for the cosmetic industry. In this thesis the choice is made to apply multiple importance sampling to the dual scattering method.

In this chapter the theoretical approach discusses the theoretical challenges involved in rendering hair models using importance sampling. Based on the literature provided in the background, there is no official approach on how to apply importance sampling to the dual scattering method. This is where this chapter kicks in. In this chapter I will elaborate on how to importance sampling the dual scattering method of Zinke et al. [?].

The theoretical challenges are as follows:

- How to apply multiple importance sampling to the dual scattering method? The multiple importance sampling algorithm that is applied to the dual scattering method has been created for the improved Marschner model. The challenge is that the dual scattering method forms a different basis to the importance sampling algorithm. It could work less optimal if the probability distribution for the importance sampling algorithm does not match accurately with the distribution of the dual scattering method. In section ?? the design decisions are explained and in section ?? the multiple importance scattering algorithm is explained in depth.
- How to measure or compare the results for multiple importance sampling? Rendering images using multiple importance sampling can look nice, or more visually pleasing when more samples are taken, but it

is important to be clear on how to evaluate rendered images. In section ?? the evaluation criteria are expressed.

- The theoretical approach to implement the multiple scattering function in the PBRT framework. Although PBRT is about the implementation of the algorithm, the framework requires an importance sampling function to be implemented in a specific way. This way has an impact on the theoretical approach.

In the following chapter Implementation, the practical challenges are discussed that occurred when implementing a physics based hair rendering algorithm.

4.1 Problem 1: How to apply multiple importance scattering to the dual scattering method?

Creating a multiple importance sampling strategy is a mathematical involved process. Except for the most simple mathematical distributions, it usually requires quite a lot of mathematical background theory in order to solve the complex integrals that follow from complex scattering distributions. This is one of the reasons why it is difficult to find an importance sampling strategy for the original Marschner model. This is also the reason why so much effort has been put in expressing the Marschner model using more simple mathematical equations. This is to make importance sampling possible as has eventually been accomplished by d’Eon et al. ??.

The current status in hair rendering can be summarized by the following two points.

- The dual scattering method provides an approximation to rendering multiple scattered hair by breaking up the computations for single scattering and multiple scattering. Single scattering resembles the original Marschner model, but the more expensive multiple scattering has been replaced by a volume scattering approximation. It provides a speed up compared to the traditional Marschner model, where extensive path tracing is needed to compute the effects of multiple scattering through the hair volume.
- The original Marschner model has been extensively analyzed and improved. One of the improvements is that the expensive root-solving process is removed. Eventually d’Eon et al. [?] improved the algorithm even further by proposing an importance sampling algorithm on top of the already improved Marschner model.

One of the observations made in d’Eon et al. [?] is that the importance sampling strategy should work for the original Marschner model, as well

as for all models that approximate the Marschner model. The observation behind this is that all approximated solutions to the Marschner model in essence are modeling the same scattering distribution. If the scattering distribution would be different, then the hair rendering would look different compared to the Marschner model. This means that the importance sampling should be able to be applied to the dual scattering method, since the dual scattering method approximates the Marschner model.

The only caveat is that the Marschner model is not fully energy conserving at the edge-cases, such as glancing angles. This problem is applicable to the dual scattering method as well. The expectation is that these edge-cases can be ignored and therefore the choice has been made that it is possible to apply the importance sampling strategy by d'Eon et al. [?] to the dual scattering method of Zinke et al. [?]. The details on the multiple importance sampling strategy are elaborated in section ??.

4.2 Problem 2: How to evaluate rendered images using multiple importance sampling?

The main goal of this thesis is to find out if multiple importance sampling applied to the dual scattering method leads to a significant reduction of noise in the rendered image. In order to come to a conclusion, evaluation criteria need to be set up. This is a challenge since there is no reference image to compare with. There are three criteria to take into account:

- Which sampling technique to use to render the reference material?
- How many samples per pixel to use for comparing rendered results?
- How to evaluate the reference image with the multiple scattered image?

The reference image will from now on be named the ground truth. In this thesis the ground truth image is rendered using the same dual scattering implementation as the multiple importance sampled results. The difference is only in the way of sampling, where the ground truth is rendered using uniform sampling. Uniform sampling can be considered a typical way of sampling when the distribution is unknown, and therefore a valid way to compare the multiple importance sampled results with. By changing only the sampling strategy, the resulting noise can be compared pixel by pixel.

The idea behind unbiased sampling strategies is that they all converge to the same result. Using this property, the expectation is that by taking a lot of samples, the result of uniform sampling and importance sampling should be equivalent. For this thesis, the ground truth is rendered up to 512 pixels per sample. This does not completely remove the noise from the ground truth, but it is sufficient to perform a comparison. Especially with

importance sampling, you should see the impact clearly for a low amount of samples. Also, taking more samples drastically increased rendering times.

At last an evaluation needs to be performed. In this thesis, the rendered results are compared for an incremental number of samples per pixel, from 1 up to 512 samples per pixel. To evaluate whether importance sampling reduces noise, two comparisons are performed.

- First of all a comparison is made by comparing the visual appearance of the image. Is the result obtained with importance sampling looking as we expect for different scenarios? Such as when the model is backlit or frontlit, or when the amount of reflection is high? This comparison is rather subjective, so there needs to be another unbiased way to measure the performance between uniform and importance sampling.
- Second of all, a comparison is performed between the variance in the rendered result. It is important to only compute the variance for the hair volume and not for the background image. This could otherwise give us a false sense of conversion if the hair model covers only a small region of the image. To prevent this from happening, only a specific rectangular region of the image is taken into account that is fully covered by the hair model. The variance is then only computed for this specific region. Figure ?? visualizes this the comparison region.

4.3 Theoretical details about importance sampling for dual scattering method

As mentioned a couple of times before in this report, d’Eon et al. [?] proposed a physically based importance strategy for the improved Marschner model (also by d’Eon et al. [?]). This strategy should work for the original Marschner model as well. In this section the theoretical details of this physically based importance sampling strategy is explained and how it is applied to the dual scattering method.

As has been explained in section ?? faster convergence can be obtained by using importance sampling over uniform sampling. The challenge with importance sampling is to sample an incoming light direction ω_i , given an outgoing direction ω_o . This is how it works in PBRT and most physics based system. The outgoing direction could be treated as the direction to the viewer or camera (at least for the first ray). The challenge is two fold: find a sampling distribution that matches the scattering behavior, and subsequently find an efficient way to sample the scattering distribution. If you are lucky, the scattering distribution is not complex and a sampling algorithm is trivial to find. If the distribution is too specific or complex, it might actually be impossible to analytically come to a sampling strategy.

The reason why the theoretical details are explained in the approach chapter is that the translation from the algorithm to code is rather challenging. The challenge comes from the way what PBRT expects an importance sampling algorithm to be provided. In general there are two situations that need to be handled when implementing an importance sampling strategy in PBRT.

- Sampling an incident (light) direction ω_i , given an outgoing (viewing) direction ω_o . The used probability distribution function should ideally be a close fit to the scattering distribution of the hair rendering model. Also the corresponding PDF should be computed, which is the likelihood of the sample being drawn.
- Compute the PDF for two given directions ω_i and ω_o . This is similar to the previous point, except that the output is now only a PDF value.

The scattering function for the Marschner model is given in equation ???. D'Eon et al. [?] simplified this equation so that the longitudinal scattering function M is only dependent on longitudinal angles, and the azimuthal scattering function N is only dependent on the azimuthal angles. Also the inclination dependent factors, such as the $\cos^2 \theta_d$ term, present in the original Marschner model are included in the M_p term. This leads to the following simplified equation:

$$S_p(\omega_i, \omega_r) = M_p(\theta_i, \theta_o) N_p(\theta_i, \theta_o, \phi) \quad (4.1)$$

4.3.1 Strategy outline

In this section the multiple importance sampling strategy for the Marschner model is explained. The strategy comes from d'Eon et al. [?] and is summarized here.

The idea behind the importance sampling strategy is to separate the sampling for the longitudinal scattering function and for the azimuthal scattering function, hence we do have multiple importance sampling. Multiple importance sampling is nothing more than a combination of two or more importance sampling strategies combined into a single multiple-importance strategy.

The basic outline of the strategy is as follows.

- Select a lobe $p \in \{R, TT, TRT\}$ dependent on the relative contribution of energy reflected. This depends on the Fresnel equation and the absorption coefficient σ .
- Given a selected lobe p , compute the longitudinal angle by sampling the longitudinal scattering function M_p .

- An outgoing direction ω_o is given, so we know the spherical angles θ_o and ϕ_o .
- Choose a random offset $h \in [-1, 1]$ along the fiber cross section.

The following subsections delve deeper into the items mentioned above.

4.3.2 Lobe selection

When importance sampling incident directions, we want to take into account the relative contribution of the different lobes in the model. For example, the R component has a stronger influence when viewing hairs from glancing angles compared to the other reflection modes. In that case you want to prefer sampling an incident direction based on the reflection component. Likewise, hair strands that are backlit do have more contribution from the TT component. It is important to take this into account, because for importance sampling we prefer to sample directions that have more impact (or weight) on the rendered result than just random sampling. With random sampling a lot of samples could be wasted, for example by taking into account the TRT component when the R reflection is much more dominant in the rendered result.

d’Eon et al. [?] proposes a way to select lobes by taking into account the attenuations through a smooth, ideally specular fiber $A_{\text{spec}}(p, h)$. The strategy is to first select a random cross-section offset $h = 2\xi_h - 1$, where $\xi_h \in [0, 1]$ is a uniformly distributed random number. Once the random cross-section offset is known, we can compute the relative contributions for each of the scattering modes. The relative contributions for the scattering modes are as follows, where 0 corresponds to R, 1 corresponds to TT and 2 corresponds to the TRT scattering mode.

$$A_{\text{spec}}(0, h) = F(\eta, \gamma_i) \quad (4.2)$$

$$A_{\text{spec}}(1, h) = (1 - F(\eta, \gamma_i))^2 T(\sigma_a, h) \quad (4.3)$$

$$A_{\text{spec}}(2, h) = (1 - F(\eta, \gamma_i))^2 F\left(\frac{1}{\eta}, \gamma_t\right) T(\sigma_a, h)^2 \quad (4.4)$$

These values describe the relative amount of energy transmitted through or reflected against hair fibers for each of the scattering modes. We do not yet take into account rough fibers, so all energy is concentrated in the specular direction. To sample a lobe these contributions should be related to the sum of all contributions to find the relative weights w_p .

$$w_R = \frac{A_{\text{spec}}(0, \gamma_i)}{\sum_p A_{\text{spec}}(p, \gamma_i)} \quad (4.5)$$

$$w_{TT} = \frac{A_{\text{spec}}(1, \gamma_i)}{\sum_p A_{\text{spec}}(p, \gamma_i)} \quad (4.6)$$

$$w_{TRT} = \frac{A_{\text{spec}}(2, \gamma_i)}{\sum_p A_{\text{spec}}(p, \gamma_i)} \quad (4.7)$$

Given the relative weights, a lobe can be sampled by taking a random uniform variable $h \in [0, 1]$ and selecting a lobe proportional to the relative attenuation of each component. Once the lobe is selected, importance sampling the longitudinal and azimuthal scattering functions becomes relatively straightforward.

The PDF of the sampling scheme is exactly analogous to evaluation of the model S in equation ??, but with the attenuations A replaced by the selection weights w_p . By replacing the attenuations with the sample weights, the PDF is always integrating to 1 irrespective of the hair color.

Finding the PDF when directions ω_o and ω_i are known

The PBRT framework requires a function to compute the PDF given two directions ω_o and ω_i . As stated in the previous section the PDF is analogous to the evaluation of the model S , but with the attenuations replaced by the selection weights w_p . The selection weights were determined before by sampling a cross section, giving us the relative contribution per scattering component and thus eventually the relative direction to be sampled.

When ω_o and ω_i are known, the PDF cannot be found by sampling from a cross section. In this case, the relative direction between the angles is used. All scattering modes could produce the difference angle, although for some scattering modes it would be more likely. In this case the process is inverse and instead of selecting a lobe given the outgoing direction, now we should take into account the likelihood of each scattering component sampling the given incident direction.

This is done by solving for roots as has been explained in the related work section ?? . The R and TT components have exactly one root and the TRT component results in one or three roots. This means that any scattering mode could have been selected that would result in the incident direction being sampled. In other words, we need to know the relative probability of a root h to be selected for each of the R, TT and TRT scattering modes, given ω_i as a sampled direction.

This is done in a similar way to sampling a lobe by using the relative attenuation for each of the roots h . Since the TRT component can have three roots, it means that this scattering mode actually has three paths to

generate the sampled direction. They are summed together to represent the attenuation for the TRT mode. The relative attenuations are then used as the sample weights when evaluating the model S to find the corresponding PDF.

4.3.3 Importance sampling the longitudinal M function

To importance sample the longitudinal scattering function M , we need a way to sample a gaussian distribution. The essence here is that we are given the outgoing direction $\omega_o = (\theta_o, \phi_o)$ and we need to find the incident longitudinal angle θ_i . The incident ϕ_i is determined in the next section (when importance sampling the azimuthal scattering function N).

There is a problem with the gaussian distribution of the Marschner model. First of all it is not energy conserving as has been explained before in section ???. Part of this can be solved by multiplying the half angle θ_h by two. Another problem is that especially at glancing angles, part of the energy falls outside of the boundary $\theta \in \{-\pi/2, \pi/2\}$, thereby losing energy. This is a problem for importance sampling. When energy is outside of the allowed range, we must somehow compensate this loss of energy in the remaining part of the gaussian distribution. This is pretty difficult to achieve, at least with the simple normalized gaussian distribution ranging from $\theta \in \{-\infty, \infty\}$.

d’Eon et al. [?] proposes a different energy-conserving longitudinal scattering function M_p . This function conservatively redistributes reflected radiance amongst directions on the sphere by employing spherical gaussian convolution. This function is energy conserving and makes sampling a lot easier. The newly used energy conserving scattering function M is as follows:

$$M_p(v, \theta_i, \theta_r) = \frac{\text{csch}(1/v)}{2v} e^{\frac{\sin(-\theta_i) \sin \theta_r}{v}} I_0\left[\frac{\cos(-\theta_i) \cos \theta_r}{v}\right] \quad (4.8)$$

where v is the longitudinal scattering variance (β_p^2) and I_0 is the modified Bessel function of the first kind. In the implementation, the Bessel function is available in the Boost library. Another option is to use the Python scientific programming library (scipy) and store the results in a lookup table.

Using the conserving scattering function, instead of the original Marschner scattering function is not an issue. This is the power of importance sampling, to be able to sample an arbitrary complex function (the marschner model) using a simpler and easier to sample function (the conserving scattering function). As long as the function is a good fit, variance will be reduced. More information about the derivation of this function and why it works can be found in the paper by d’Eon et al. [?].

Sampling a spherical gaussian can be done with two random numbers ξ_1 and ξ_2 each within $[0, 1]$. Using the Box-Muller transform gives us a sample

for the normalized spherical gaussian function. The computation requires a lot of mathematics, which is explained in the paper by d'Eon et al [?]. The strategy to find the sampled θ_i is as follows.

$$u(\xi_1) = v \log \left(e^{1/v} - 2\xi_1 \sinh \frac{1}{v} \right) \quad (4.9)$$

$$\theta_{\text{cone},p} = -\theta_i + \alpha_p \quad (4.10)$$

$$\theta' = \frac{\pi}{2} - \theta_{\text{cone}} \quad (4.11)$$

$$\begin{aligned} \theta_i(\xi_1, \xi_2, v, \theta_{\text{cone}}) = \\ \arcsin(u(\xi_1) \cos \theta' + \sqrt{1 - u(\xi_1)^2} \cos(2\pi\xi_2) \sin \theta') \end{aligned} \quad (4.12)$$

The probability of sampling this value is equal to the longitudinal scattering function M , assuming that the scattering distribution is energy conserving and integrates to 1, so the PDF is $M \cos^2 \theta_i$.

4.3.4 Importance sampling the azimuthal N function

Importance sampling the azimuthal N function is trivial if we know the selected lobe p and the offset h from the scattering cross section (see section ??). In that case it is a matter of evaluating the relative change in azimuth $\Phi(p, h)$ (see equation ??). To find the incident azimuth direction ϕ_i .

$$\phi_i^{\text{smooth}} = \phi_o + \Phi(p, h) \quad (4.13)$$

For rough fibers, the relative change in azimuth is offset by a gaussian distributed random variable g with standard deviation 1, multiplied by the width β_p corresponding to the selected lobe p . Sampling a gaussian distribution can be performed using the Box-Muller transform as explained in the previous subsection. The direction ϕ_i to be importance sampled can then be found as follows:

$$\phi_i = \phi_o + \Phi(p, h) + g|\beta_p| \quad (4.14)$$

Take into account that the spherical direction wraps around in the range $\{-\pi, \pi\}$.

Chapter 5

Implementation

Aside from the theoretical challenges, there are also practical challenges in implementing physics based rendering algorithms, especially when dealing with hair models that often consist of hundreds of thousands of hair fibers. These hair fibers which are huge in amount, are also very thin, especially compared to the projected area on a pixels.

This chapter deals with the implementation details to implement the importance sampling strategy for the dual scattering method. The theoretical approach is described in the previous chapter. The practical challenges can be summarized as follows.

- Framework to render physics based algorithms in combination with the representation of hair strands.
- How to represent the voxel grid
- How to deal with memory and speed requirements

The next sections will address these challenges on how these are tackled in this thesis.

5.1 Problem 1: Which framework to use to render physics based?

The first challenge is to find a proper ray tracing framework, which offers the possibility to write your own BSDF materials. There is a lot of choice between different renderers, but one that is extremely useful is PBRT. PBRT stands for Physics Based Ray Tracing and is used a lot to explore and experiment with physics based raytracing algorithms. It is developed by Matt Pharr et al. and freely available, open source from their website] and de-

⁰<http://www.pbrt.org>

scribed in their corresponding book [?]. The rendered images in this thesis are created using PBRT.

Within PBRT, it is possible to create your own BSDF's. A BSDF is usually applied to a surface as has been discussed in section ??, but for hair rendering we are using curve representations. The curve variant of a BSDF is called a "bidirectional curve scattering distribution function", abbreviated as BCSDF. PBRT provides a generic solution using BxDF classes in which it is possible to represent BCSDF materials.

The representation of hair fibers can be done in a variety of ways as has been discussed in section ???. The possibilities were to represent them as connected triangle strips, cylindrical primitives, trigonal prisms or ribbons. Within PBRT hair fibers are specified as curves, which are a sequence of cubic Bezier splines represented by 3D points. For every cubic spline segment, the width is specified. PBRT has support for three types of curves: flat, cylindrical or ribbons. Flat corresponds to triangle strips, and the other curve types are self-explanatory. For the rendered results in this thesis cylindrical curves are used, which correspond to the most realistic representation.

5.2 Problem 2: How to write code for voxel grid?

Voxel grid implementations are not that trivial as they might seem at first. Yes, you can use a three-dimensional array, but this is not efficient for rendering purposes. One of the problems with voxel grids are the hard edges they produce in the renderings. Making the voxel cells very small leads to memory issues, or even precision issues because of their small size.

The dual scattering approximation requires a way to efficiently approximate the global multiple scattering. The global multiple scattering can be computed in a number of ways. Zinke et al. [?] proposed the following methods:

- Ray-shooting: with ray-shooting a shadow ray is shot through the hair volume, calculating the number of intersections with the hair fibers, together with the angle the ray makes with the direction of the fiber. In this way the density can be approximated using the formulas as explained in the background theory on dual scattering ???. This strategy is the simplest but most expensive one to compute, because of the number of intersections being computed. Also this strategy is a little tedious to implement in PBRT. The reason for this, is that PBRT is build on the fact that every intersection is a material intersection. Tracing a shadow ray through a volume, just to count the number of intersections, is actually something non-physical and therefore not supported out of the box.
- Forward-scattering maps: with forward scattering maps a two-pass ap-

proach is made. First a voxel grid is precomputed in which every voxel cell contains containing the average values for the forward transmittance T_f and forward longitudinal spread σ_f^2 . During the second pass, linear interpolation is used to integrate along the ray through the voxel grid.

- GPU rendering using deep-opacity maps: this solution is very fast, but contains some tricks using deep opacity maps and a preprocessing step where depth maps are calculated. This thesis primarily focuses on physics based approaches and therefore didn't went in this direction.

The most plausible approach for this thesis is the second option: a two pass approach in which a voxel grid is precomputed. From the explanation it follows that the voxel grid should contain a possibility to do linear interpolation along a path. Also from a practical perspective, the voxel grid should not consume too much memory. For this thesis, an open-source software package OpenVdb is used ¹. OpenVDB is developed at Dreamworks Animation and is especially suited to be used when rendering volumetric effects. It is able to store sparse voxel grids without consuming much memory and with support for different ways of sampling: point, box and quadratic sampling that correspond to averaging out the sample based on the value in the neighboring voxel cells.

In this thesis the preprocessing step works as follows. For every hair fiber that is represented as a curve, loop through the curve by a specific offset, and for each position, find the corresponding voxel cell. This voxel cell is incremented with the length of the curve segment. In this way, all voxel cells, contain the relative lengths that have been traversed in that voxel cell. These lengths do not necessarily correspond to absolute density values, but are relative density values. When rendering a hair model, the density can be scaled to adjust the rendered result. In this thesis I decided not to store the orientations per voxel cell. This is a simplification. The reason I chose for this simplification is that hair fibers are usually similar in orientation in their local neighbourhood. Also it makes the voxel generation process less complex. For more realistic behavior that is applicable to more complex hair models, the orientations of the fibers should be taken into account.

In the second pass the hair model is rendered. If the ray intersects with the first hair fiber at the surface of the hair volume, a lookup is performed in the voxel grid. The voxel grid, based on 3D location and orientation, is used to traverse through all voxel cells along this "shadow" ray, integrating the results using box sampling. The choice for box sampling is arbitrarily, because the results were nice enough. It could be worth to do further research on choosing different sampling algorithms and on determining the best possible

¹<http://www.openvdb.org>

voxel size for different types of hair models such as straight hair versus curvy hair.

One experienced side effect of integrating the density through the voxel grid, is for intersections at the hair volume boundary. The first fiber hit is not in shadow. This is evident, since it's located at the boundary edge. However, looking up the density value gives an indication that the fiber is in shadow. This is solved by tracing an additional shadow ray to the light source to decide whether it has a clear path to the light source. If that is the case, no lookup is performed and it is assumed the hair fiber is fully lit without strands covering it.

5.3 Problem 3: How to deal with memory and speed requirements?

Rendering hair models is very slow. The system settings that I used for rendering is an Intel Core i7-3770K processor with 8 GB of RAM. Rendering times for 512 samples per pixel could lead up to hours, and the curvy hair model was not possible to render at all, because it consumed more memory than the 8 GB of available RAM. The reason why memory usage is so high, is because some bookkeeping structure is set up by PBRT to improve ray intersections.

One of the tricks that is performed in this thesis is to do the preprocessing step, computing the voxel grid, with the full hair model. The rendering step is then performed by using half the hair model, where the 'cutted' half of the hair model is viewing away from the camera. In this way, it is not visible that actually half the hair model is used during rendering. This doesn

Chapter 6

Results

In this chapter results are presented. First the dual scattering model is presented in a variety of settings. Then the effect of the average forward and backward density factors are shown, followed by a numerical analysis by using variance to compare the noise between uniform sampling and importance sampling.

6.1 Dual scattering in real-world situations

Usually when rendering, multiple importance sampling is performed to take samples from the light source and samples determined by the material. If the light source would be small, then taking samples determined by the light source will usually always give better results than using importance sampling from the material. To be able to see the effect of importance sampling for the hair material, we thus must make sure that we have a globally illuminated scene, to see the difference between importance sampling and uniform sampling.

The results presented in this section compares results from uniform sampling with results from importance sampling. The models are placed in a global illuminated scene, where global illumination is accomplished by using HDR spherical environment maps.

In figure ?? results are presented for a daytime setting between office buildings. Here the daylight is abundant but in general no direct sunlight shining on the hair models.

In figure ?? results are presented for an outside square in Venice.

In figure ?? results are presented for a dark setting. The scene is a subway station platform, where the lights in the station provide the direct lighting (and reflections from the walls and floor the indirect lighting). The models likewise appear very dark. Since the absorption inside blonde hair is relatively low, this model still clearly shows resemblance to a blond hair color, while the brown and black hair models are actually very dark.

Observations

By observing the results uniform sampling seems to appear a bit more noisy, but not in such amount that you could call it a significant difference. Also it shows that the blonde hair model is the brightest of all. Especially in the subway setting (figure ??) the blond hair still resembles blond hair, while the other two hair models are very dark, almost not being able to see a distinction between black and brown hair. Black hair in general has a minimal amount of global multiple scattering, because the absorption coefficient is so high. Light passes through this hair in minimal amount, therefore appearing black. The only effect visible in black hair is a direct reflection according to the reflection (R) component.

6.2 Variation of the forward and backward density factors

Figure ?? shows the effect of changing average forward and backward density values d_f and d_b . When the density factors are zero, the appearance boils down to single scattering. This is because a density factor of zero indicates that no global multiple scattered light reaches the shading points (refer to figure ?? for a visual representation).

The results show that the global multiple scattering component is very important to the appearance of blond hair models. To a lesser extend, but still important, to brown hair models. For black hair models the global multiple scattering is hardly visible. This is in accordance to what is expected, because light colored hair types have by definition less attenuation when light propagates through hair fibers. Therefore there is more global multiple scattering through the hair volume.

6.3 Theoretical Results

In this section theoretical results are obtained based on measured data. As discussed in the approach the variance is used to quantify the error compared with the ground truth. By varying the samples per pixel an insight is gathered in the convergence rate and the efficiency of the importance sampling strategy compared with uniform sampling. Figure ?? displays a rendering of a curly brunette with an outline indicating the region of interest for variance computation. In subsequent sections, this region is used for variance computation. This is done so that the background image has no effect on the variance computation.

Pixels per sample	Uniform Sampling	Importance Sampling
1	-	-
2	0.002931	0.001524
4	0.001058	0.000516
8	0.000443	0.000209
16	0.000192	0.000084
32	0.000113	0.000045
64	0.000044	0.000019
128	0.000025	0.000009
256	0.000102	0.000083
512	0.000003	0.000001

Table 6.1: The relative variance for increasing samples per pixel. For example, when comparing uniform sampling for 16 samples per pixel, the distance compared with 8 samples per pixel is 0.000192 (expressed in variance). If the relative variance is 0, then this indicates the doubling of samples per pixel did not cause a change with the previous image (and thus the result is converged). For 1 sample per pixel there are no results, since there is no previous image to compare with.

6.3.1 Relative variance for increasing samples per pixel

Table ?? contains measurement data for relative variance as the samples per pixel increases. The samples per pixel are doubled at every step starting from 1 sample per pixel up to 512 samples per pixel. At every step the distance is computed (as a variance metric) between the current image and the result from the previous step (where half the samples per pixel are used).

The relative variance suggests that for both uniform and importance sampling, at 512 samples per pixel there is hardly any change compared to 256 samples per pixel. This indicates that convergence has been reached. The rate of convergence shows that for the first few increments of samples per pixel, the result is changing considerably. This is what we expect: as the samples per pixel increases, we get closer to convergence and thus expect a lower and lower relative variance. Figure ?? shows partial regions of the renderings. As the samples per pixel increases, the result becomes much more pleasant to watch at.

With this data it is difficult to do a hard comparison. In the next section the images are compared with the ground truth, giving us more insight in the efficiency of both sampling strategies related to each other.

Pixels per sample	Uniform sampling	Importance Sampling (compared with uniform)
1	0.019034	0.010120
2	0.009072	0.004829
4	0.004352	0.002440
8	0.001842	0.001119
16	0.001004	0.000760
32	0.000357	0.000415
64	0.000156	0.000321
128	0.000054	0.000282
256	0.000017	0.000264
512	0.000000	0.000259

Table 6.2: These measurements quantify the error compared with the ground truth. The ground truth is taken as the uniform sampling variant at 512 samples per pixel. Therefore the error for uniform sampling is 0 at 512 samples per pixel, because they are identical. It gives an insight in the convergence rate, together with the graph shown in figure ??.

6.3.2 Variance compared to ground truth

One of the most important requirements for an importance sampling strategy is that it converges to the same result as uniform sampling. As discussed in the previous section it appears that the uniform sampling renderings are converged at 512 samples per pixel, so we will use the result of 512 samples per pixel as the ground truth. Importance sampling results can then be compared with the ground truth by computing the distance with the ground truth (expressed as the average variance per pixel).

Figure ?? shows a side by side comparison of 512 samples per pixel, assuming that both images have been converged. The importance sampled version is a bit darker compared to the uniform sampling version. This is not expected. It could indicate that the probability density function is not integrating to 1 for all configurations. A possible explanation could come from the work by d’Eon et al. [?] in which they explain that the Marschner model is not energy conserving. Doubling the half angles as we have done in this thesis solves part of the problem. Still deflection near grazing angles moves considerably into angles outside the range of possible angles, thereby losing energy.

Figure ?? shows an even more detailed side by side comparison. Here a region is displayed to better see the noise for various samples per pixel for both uniform and importance sampling. Table ?? shows the variance of different renderings compared to the ground truth and figure ?? displays the measurement data in a plot.

There are a couple of observations to be made:

- Initially the variance is decreasing rapidly. When the samples per pixel become sufficiently large, then the improvement in variance becomes less and less, both for uniform sampling and for importance sampling. This can also be seen from the side by side comparison in figure ?? . This is expected as you come closer to the converged result.
- Importance sampling performs better for small samples per pixel. This is expected, because for smaller amount of samples importance sampling can do a better job at finding important samples. Eventually uniform sampling catches up as they both reach the converged image.
- Importance sampling variance is crossing the uniform sampling variance at 32 samples per pixel. From that point on uniform sampling is closer to ground truth than importance sampling. This has to do with the fact that this importance sampling strategy does not converge to the ground truth (for reasons mentioned above).

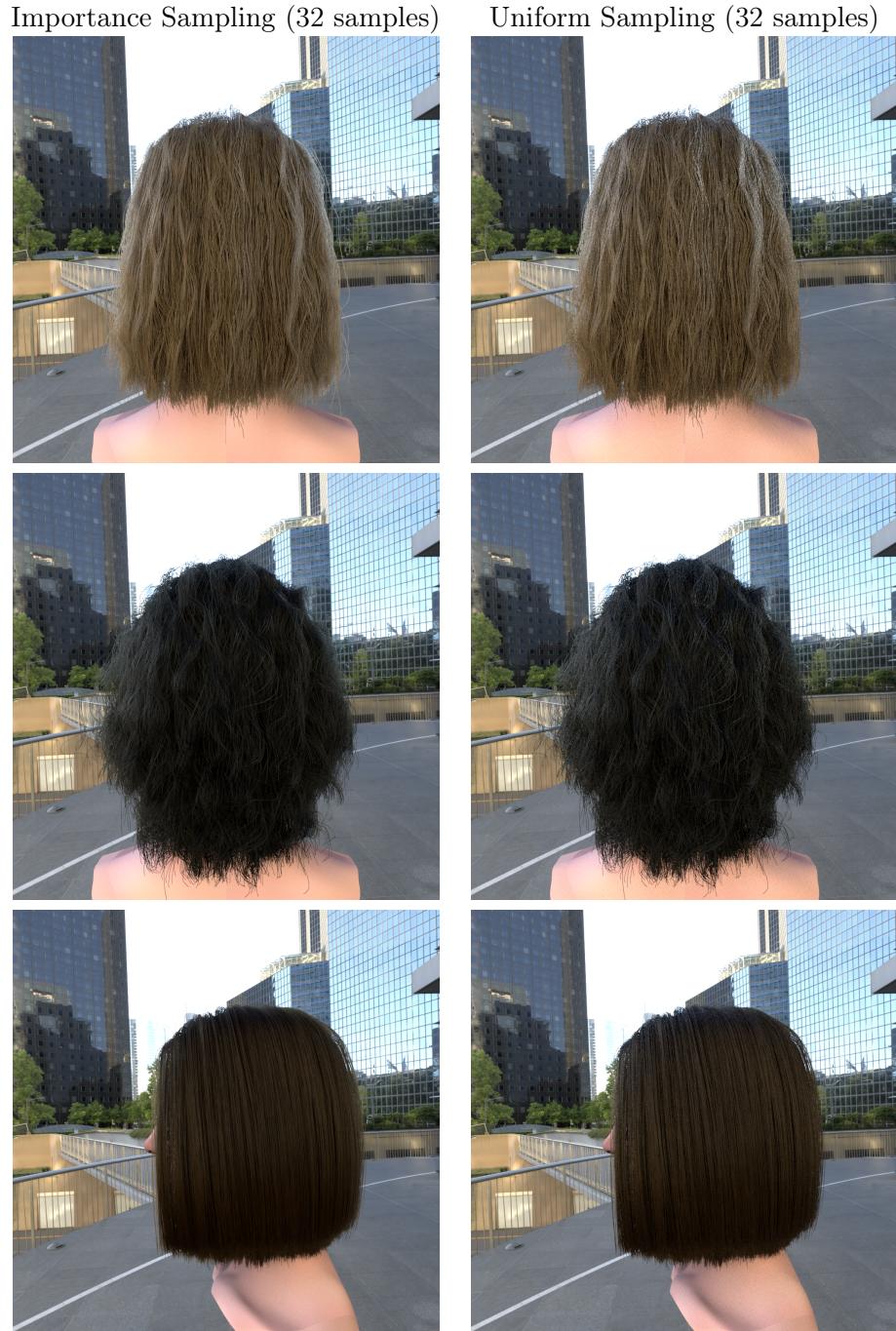


Figure 6.1: Comparison between importance sampling and uniform sampling in an office square setting. The images are rendered using 32 samples per pixel.

Importance Sampling (32 samples)



Uniform Sampling (32 samples)



Figure 6.2: Comparison between importance sampling and uniform sampling in an office square setting. The samples are rendered using 32 samples per pixel.

Importance Sampling (32 samples)



Uniform Sampling (32 samples)

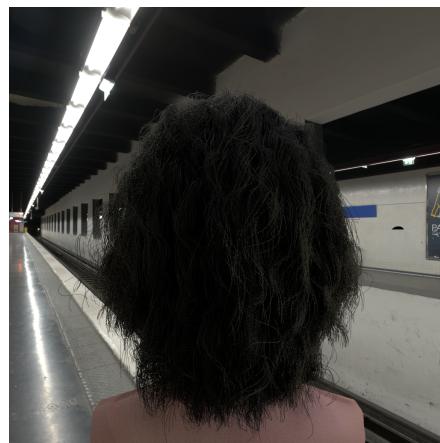


Figure 6.3: Comparison between importance sampling and uniform sampling in an office square setting. The samples are rendered using 32 samples per pixel.

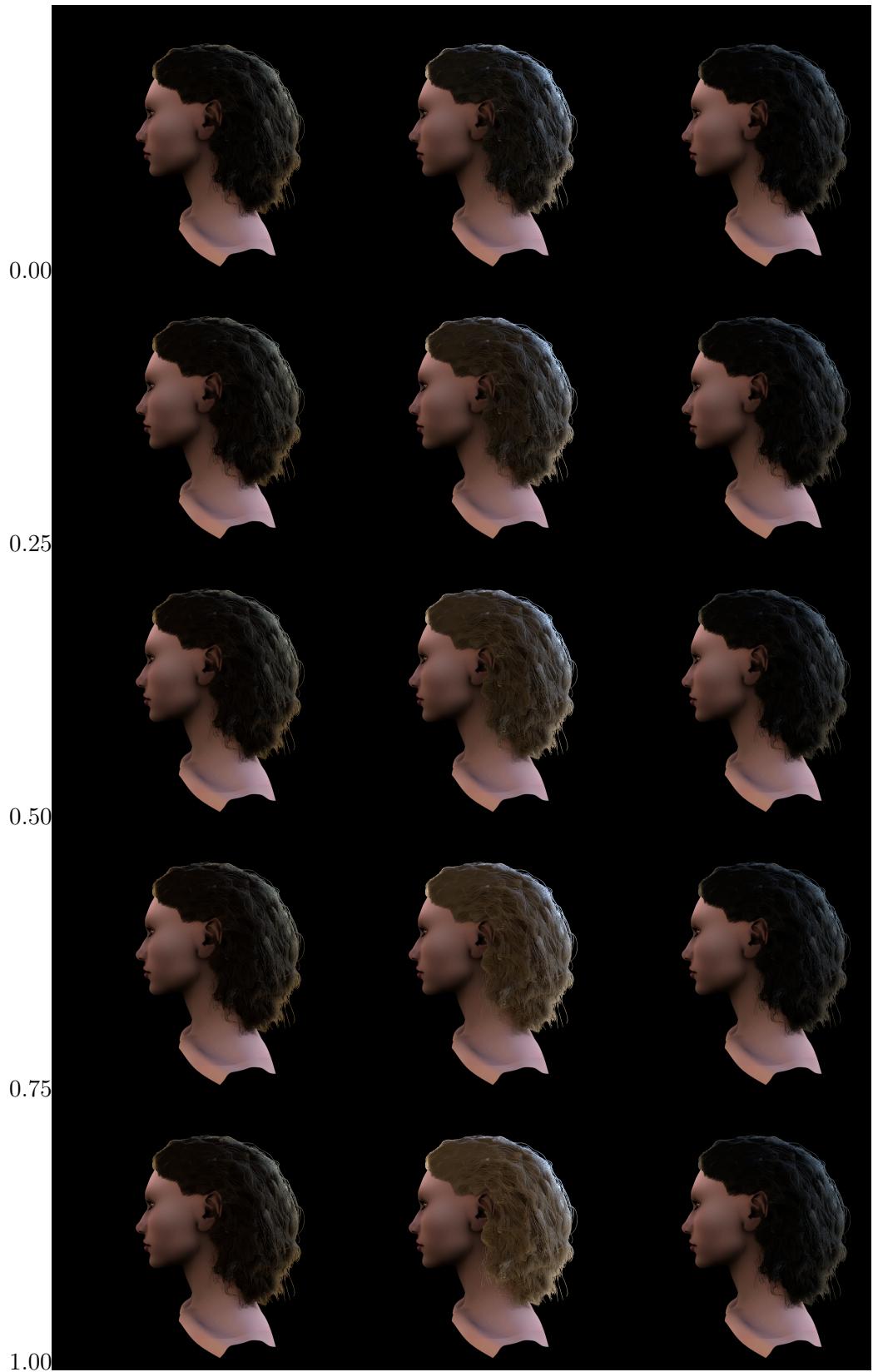


Figure 6.4: A variation of the density factors d_f and d_b , corresponding to the amount of global multiple scattering in the hair model. A value of 0 indicates the absence of global multiple scattering and 1 the full presence of global multiple scattering. Usually the value is set to a constant value 0.7.

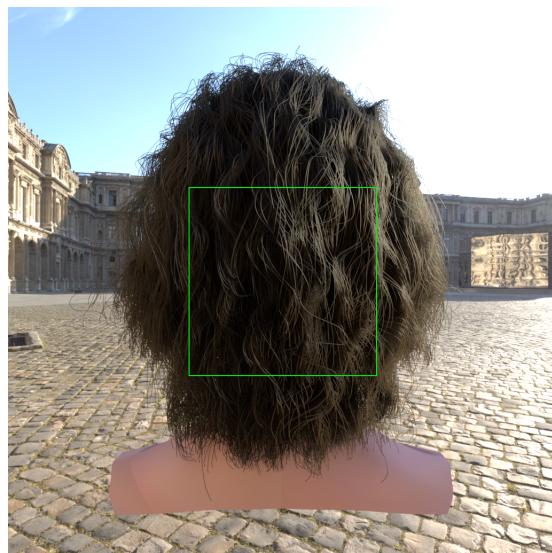


Figure 6.5: This figure displays an outline representing the region of interest for subsequent comparison of rendered results. The variance is computed for the area contained by the outline. This is done so that the background, and nonrelated items have no influence on the variance computation.



Figure 6.6: Comparison of the image quality between uniform sampling and importance sampling for various pixels per sample, indicated in the bottom left corner. As the pixels per sample increase the noise of the image becomes smaller.



Figure 6.7: A side by side comparison of converged images for uniform and importance sampling, both rendered at 512 samples per pixel. Both renderings should converge to the same result. Although the result looks similar, the uniform sampling seems slightly brighter.

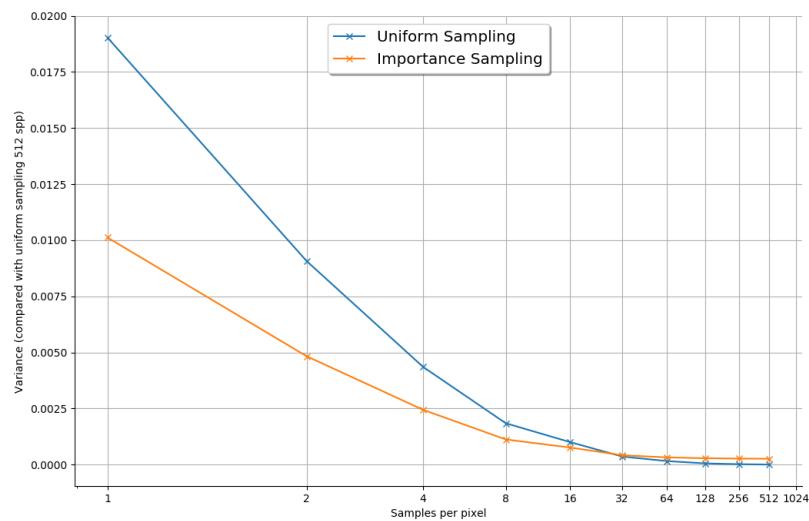


Figure 6.8: Performance analysis between uniform sampling and importance sampling in terms of variance. The ground truth is taken to be uniform samples at 512 samples per pixel. The data corresponds to data in table ??.

Chapter 7

Conclusion

The goal of this report is to evaluate the multiple importance sampling strategy proposed by d'Eon et al. [?] for the dual scattering method proposed by Zinke et al. [?]. In this report a couple of challenges were mentioned.

First of all, the dual scattering method is based on the original Marschner model which is not energy conserving. This makes it difficult to find an importance sampling strategy. Several improvements have been proposed over time to fix the issues with the Marschner model, making it possible to importance sample the Marschner model. Since the dual scattering method in this report is based on the original Marschner model, it is not an accurate fit. This could be one of the reasons why the rendered results look a tad darker when comparing the uniform sampled with multiple importance sampled rendering results.

Second of all, ... How to evaluate rendered images using multiple importance sampling? Theoretical Approach: How to prepare for PBRT.

Implementation Problem 1: Which framework to use to render physics based? Problem 2: How to write code for voxel grid? Problem 3: How to deal with memory and speed requirements?

It turns out that importance sampling leads to more efficient renderings, but in such a small amount that it is hardly noticeable. One of the reasons that I can think of is the fact that the dual scattering method smoothens out the global illumination, essentially meaning that light is coming from almost anywhere around the hemisphere (for outside fibers) or sphere (for inside fibers). Sampling specific directions that are likely to contribute a lot to the Marschner model are not so significant anymore to the dual scattering method. This suggests that more work can be done in finding a way to importance sample the dual scattering method.

One of the problematic issues when using the Marschner model in combination with the dual scattering method is the fact that the parameters used to change the appearance of the model are not intuitive. Also the root finding process is quite complex and d'Eon et al. [?] proposed a simpler

and energy conserving model which produce similar results. Also Sadeghi et al. [?] proposed a less physically based artist friendly hair shading system, where physical properties are replaced by intuitive controls to color the hair, especially useful for artists for animation moview.

The simplifications used, such as assuming that the orientations of hair strands are the same when determining the global multiple scattering contribution, leads to similar appearance for neighboring hair strands. These appearances seem unnatural because of the washed out look. More work could be done by taking into account the different orientations for hair fibers in the local neighborhood. Also the use of the voxel grid could lead to less optimal results, since all positions in a voxel cell are using the same data (smoothed by trilinear filtering). Experimentation with voxel cell sizes could be performed, but a better idea is to use deep opacity maps (as originally proposed by Zinke et al. [?]).

Also it would be beneficial to try to apply dual scattering method to the solution proposed by d'Eon et al. [?]. In that way there is a proven 100 percent energy conserving model that might lead to better results, so that importance and uniform sampling are converging to exactly the same result.