Karlsruhe Institute of Technology

Importance Sampling in Many Lights Trees

Bachelor's Thesis of

Beini Ma

at the Department of Informatics
Computer Graphics Group

Reviewer:	Prof. A
Second reviewer:	Prof. B
Advisor:	M.Sc. Alisa Jung
Second advisor:	M.Sc. D

23. April 2018 – 23. August 2018

Karlsruher Institut für Technologie
Fakultät für Informatik
Postfach 6980
76128 Karlsruhe

I declare that I have developed and written the enclosed thesis completely by myself, and have not used sources or means without declaration in the text.

Karlsruhe, 23. August 2018

. . . . . .

(Beini Ma)

Abstract

English abstract.

Zusammenfassung

Deutsche Zusammenfassung

Contents

List of Figures

List of Tables

1 Introduction

1.1 Problem/Motivation

Path tracing is one of the most important rendering techniques when creating highly realistic pictures. It allows us to render the scene much closer to reality compared to typical scanline rendering methods at the cost of more computations. In situations where the images can be rendered ahead of time, such as for visual effects or films, we can take advantage of the better results of ray tracing. Then again, ray tracing is not yet useful for real-time applications like video games where the rendering speed is critical. But even regarding ray tracing, we cannot completely ignore the rendering time. Too long rendering times are becoming a problem in scenes with many lights. For instance, a scene that consists of a big city with skyscrapers at night could have hundreds or thousands of lights that could potentially all affect a single point in the scene. Lighting methods that calculate the incident lighting of a point for every single light in the scene would be too slow to deal with these kinds of scenes since every ray to the camera could potentially trace multiple points that all need to be lighted.

There are sampling approaches that try to limit the time required to render these scenes with a big amount of lights. For instance, we could say that the probability of a point of the scene sampling a certain light is only dependent on the emission power of said light. We would make a distribution that only takes into account the emission power of the lights. To light a specific point we would then sample a single light according to the distribution function we built earlier. Obviously there are a lot of problems with this approach. An area light source or a spotlight could be facing towards a completely different direction and may not have any effect on the point. Or the light source could be potentially too far away to have a noticeable effect on the point. This sampling technique asserts a fast sampling speed but can lead to very noisy images that we are trying to avoid.

For this bachelor's thesis we will introduce a light sampling technique that improves the rendering speed without making the rendered image too noisy.

1.2 Content

2 Preliminaries

2.1 Probability Theory Basics

In this section we will be discussing basic ideas and define certain terms from the probability theory. We will assume that the reader is already familiar with most of the concepts and therefore will only give a short introduction. If the reader struggles following the key parts of this section, he is heavily advised to read more extensive literature about this subject. We suggest E. T. Jaynes *Probability Theory: The Logic of Science* for this matter. [Jay03]

2.1.1 Random Variable

A random variable X is a variable whose values are numerical outcomes chosen by a random process. There are discrete random variables, which can only take a countable set of possible outcomes and continuous random variables with an uncountable number of possible results. For instance, flipping a coin would be a random variable drawn from a discrete domain which can only result to heads or tails, while sampling a random direction over a unit sphere can produce infinite different directions. In rendering and particularly in path tracing, we are often sampling certain directions or light sources in order to illuminate the scene, therefore we will be handling both discrete and continuous random variables, albeit with the latter in the most cases.

The so-called canonical uniform random variable ξ is a special continuous random variable that is especially important for us. Every interval in its domain $[0, 1)$ with equal length are assigned the same probability. This random variable makes it very easy to generate samples from arbitrary distributions. For example, if we would need to sample a direction to estimate the incident lighting on a point, we could draw two samples from ξ and scale these two values with appropriate transformations so they reflect the polar coordinates of direction to sample.

2.1.2 Probability Density Function

For continuous random variables, probability density functions (PDF) illustrate how the possible outcomes of the random experiment are distributed across the domain. They

must be nonnegative and integrate to 1 over the domain. $p : D \longrightarrow \mathbb{R}$ is a PDF when

$$\int_D p(x)\mathrm{d}x = 1. \tag{2.1}$$

Integrating over a certain interval $[a, b]$ gives the possibility that the random experiment returns a result that lies inside of given interval:

$$\int_a^b p(x)\mathrm{d}x = P(x \in [a, b]) \tag{2.2}$$

It is evident, that $P(x \in [a, a]) = 0$ which reflects the fundamental idea of continuous random variables: The possibility of getting a sample that exactly equals a certain number is zero. Therefore, PDFs are only meaningful when regarded over a interval and not over a single point.

2.1.3 Expected Values and Variance

As the name already indicates, the expected value $E_p[f(x)]$ of a function f and a distribution p specifies the average value of the function after getting a large amount of samples according to the distribution function $p(x)$. Over a certain domain D, the expected value is defined as

$$E[f(x)] = \int_D f(x)p(x)\mathrm{d}x. \tag{2.3}$$

The variance defines a measure that illustrates the distance between the actual sample values and their average value. Formally, it is defined by the expectation of the squared deviation of the function from its expected value:

$$V[f(x)] = E[(f(x) - E[f(x)])^2] \tag{2.4}$$

When we talk about Monto Carlo Intergration later, the variance is a strong indicator of the quality of the PDF we chose. The main part of this thesis will be to minimize the variance of light sampling methods.

2.1.4 Error and Bias

For an estimator F_N, the parameter to be estimated I and a sample x the error $e(x)$ is defined as

$$e(x) = F_N(x) - I. \tag{2.5}$$

Since the error is dependent on the certain sample we took, we will also introduce bias. The bias b of an estimator is the expected value of the error:

$$b(F_N) = E[F_N - I]. \tag{2.6}$$

An estimator F_N is unbiased, if $b(F_N) = 0$ for all sample sizes N. Informally, it means that the estimator is going to return the correct value on average. In the next section, we will introduce an unbiased estimator, the Monte Carlo estimator.

2.2 Monte Carlo Integration

When generating an image using path tracing, we will be dealing with integrals and our main task will be to estimate the values of these integrals. Since they are almost never available in closed form, like the incident lighting of a certain point that theoretically requires infinite number of rays traced over infinite dimensions, analytical integration methods do not work. Instead, we have to use numerical integration methods give an appropriate estimation for these integrals. One of the most powerful tools we have in this regard is the Monto Carlo integration. We will be discussing the advantages of Monte Carlo integration, as well as its constraints and mechanisms how we can deal with these limits.

Using random sampling methods, we want to evaluate the integral

$$I = \int_a^b f(x)d(x) \tag{2.7}$$

with Monte Carlo integration. Different to Las Vegas algorithms, Monte Carlo integration has a non-deterministic approach. Every iteration of the algorithm provides a different outcome and will only be an approximation of the actual integral. Imagine that we want to integrate a function $f : D \rightarrow \mathbb{R}$. The Monte Carlo estimator states that with samples of uniform random variables $Xi \in [a, b]$ and number of samples N the expected value $E[F_N]$ of the estimator

$$F_N = \frac{b-a}{N} \sum_{i=1}^{N} f(X_i) \tag{2.8}$$

is equal to the integral, since:

$$
\begin{aligned}
E[F_N] &= E\left[\frac{b-a}{N}\sum_{i=1}^{N} f(X_i)\right] \\
&= \frac{b-a}{N}\sum_{i=1}^{N} E[f(X_i)] \\
&= \frac{b-a}{N}\sum_{i=1}^{N} \int_a^b f(x)p(x)dx \\
&= \frac{1}{N}\sum_{i=1}^{N} \int_a^b f(x)dx \\
&= \int_a^b f(x)d(x)
\end{aligned}
\tag{2.9}
$$

If we use a PDF $p(x)$ instead of an uniform distribution, the estimator

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

is used to approximate the integral. Being able to use arbitrary PDFs is essential for solving the light transport problem and the importance of choosing a good PDF $p(x)$ will be explained in the next section.

The Monto Carlo estimator is unbiased, because

$$
\begin{aligned}
b(F_N) &= E\left[\frac{1}{N}\sum_{i=1}^{N} \frac{f(X_i)}{p(X_i)} - I\right] \\
&= E\left[\frac{1}{N}\sum_{i=1}^{N} \frac{f(X_i)}{p(X_i)}\right] - E[I] \\
&= E\left[\frac{f(X)}{p(X)} - I\right] \\
&= I - I = 0.
\end{aligned}
\tag{2.11}
$$

While standard quadrature techniques converge faster than the Monte Carlo integration in low dimensions, the Monte Carlo integration is the only integration method that allows us to deal with higher dimensions of the integrand, since it's convergence rate is independent of the dimension. In fact, standard numerical integration techniques do not work very well on high-dimensional domains since their performance becomes exponentially worse as the dimension of the integral increases. Later, we will explain why the light

transport problem of the path tracing algorithm is theoretically an infinite-dimensional problem and therefore we will estimating the integrals in our light transport equations with Monte Carlo integration at a convergence rate of $O(\sqrt{N})$. [Vea97]

2.3 Importance Sampling

As we have mentioned earlier, the Monte Carlo estimator allows us to use any distribution $p(x)$ to get the samples from. Abusing the fact, that the Monte Carlo estimator

$$F_N = \frac{1}{N} \sum_{i=1}^{N} \frac{f(X_i)}{p(X_i)} \tag{2.12}$$

converges faster if we choose a sampling distribution $p(x)$ that is roughly proportional to the function $f(x)$, this will be our main variance reduction method. Suppose, we could use any distribution $p(x)$ to pick our samples from. We would then choose a distribution $p(x) = cf(x)$, basically a distribution proportional to the function of the integrand. Then, after each estimation, the value we would have calculated would be

$$\frac{f(X_i)}{p(X_i)} = \frac{1}{c} = \int f(x)dx. \tag{2.13}$$

We would be getting a constant value and since the Monte Carlo estimator is unbiased, every single of our estimates would exactly return the value of the integral and our variance would be zero. Obviously, that does not work, since it requires us to know the value of the integral $\int f(x)dx$ before, and then using Monte Carlo techniques to evaluate the integral would be pointless. Nonetheless, if we are able to find a distribution $p(x)$ that has a similar shape to $f(x)$, we would be archive a faster convergence rate. To put it in relation to path tracing, assume we want to calculate the incident lighting for a given point on a diffuse surface. Imagine, the light sources of the scene are all focused on a certain direction of the point. It would make a lot of sense to mainly sample the directions that are similar to the vectors pointing to the light sources, since their contribution will be the biggest. In this case, we would want to use a distribution that is similar to the light source distribution in the scene given the position of the point.

2.4 Multiple Importance Sampling

We have been discussing how to deal with integrals of the form $\int f(x)dx$ with Monte Carlo techniques. While path tracing, we are mainly faced with situation where we have to evaluate integrals that consists of a product of two functions. In fact, the main function

we need to estimate for the direct lighting given a certain point and the outgoing direction is in that form:

$$L_o(p, \omega_o) = \int_{\varsigma^2} f(p, \omega_o, \omega_i) L_d(p, \omega_i) |cos\theta_i| d\omega_i \tag{2.14}$$

$f(p, \omega_o, \omega_i)$ describes the reflectance of the surface at given point p, and the ingoing and outgoing directions ω_i and ω_o. $L_d(p, \omega_i)$ specifies the incident lighting at given point and direction ω_i. It is apparent, that using a single distribution function in every situation will not yield the optimal results. A specular surface would only reflect the light in very specific angles and in these cases, having a distribution, that has a similar form of $f(p, \omega_o, \omega_i)$ would be preferable. On the other hand, if the surface was diffuse, we would obtain better results, when using a distribution that has a similar shape of that of the incident lighting.

The solution to this dilemma is multiple importance sampling. When using multiple importance sampling, we will draw samples from multiple distributions to sample the given integral. The idea is that, although we do not know which PDF will have a similar shape to the integrand, we hope that one of the chosen distributions match it fairly well. If we want to evaluate the integral $\int f(x)g(x)dx$ and p_f and p_g are our distribution functions, then the Monte Carlo estimateor with multiple importance sampling is

$$\frac{1}{n_f} \sum_{i=1}^{n_f} \frac{f(X_i)g(X_i)w_f(X_i)}{p_f(X_i)} + \frac{1}{n_g} \sum_{j=1}^{n_g} \frac{f(X_j)g(X_j)w_g(X_j)}{p_g(X_j)}, \tag{2.15}$$

where n_f and n_g are the number of samples taken from their respective distribution function and w_f and w_g are weighting functions so the expected value of the estimator still matches the value of the integral. Multiple importance sampling is able to significantly lower the variance because a good weighting function should dampen contributions with low PDFs. We will present two weighting heuristics, the *balance heuristic* and the *power heuristic*.

A good way to reduce variance is the balance heuristic:

$$w_s(x) = \frac{n_s p_s(x)}{\sum_i n_i p_i(x)} \tag{2.16}$$

The power heuristic with exponent $\beta = 2$ is often a better heuristic to reduce variance:

$$w_s(x) = \frac{(n_s p_s(x))^2}{\sum_i (n_i p_i(x))^2} \tag{2.17}$$

For a more detailed explanation of the weighting heuristics, we recommend "Robust Monte Carlo Methods for Light Transport Simulation" by Veach to the reader. [Vea97]

2.5 Axis-Aligned Bounding Box

TODO: image

2.6 Bounding Volume Hierarchies

Acceleration data structures are mandatory for path tracers. They reduce the amount of ray-primitive intersection tests can be reduced to logarithmic in the number of object. The basic idea of those data structures is to partition the primitives into sets and order those in a hierarchy, so only specific sets of primitives need to be tested for intersections. The two main approaches, when dividing the primitives, is to either split the room among the space or to choose a particular number of primitives and wrap a bounding box around these primitives. The acceleration data structure used in PBRT is the bounding volume hierarchy (BVH). It uses the latter approach.

An example to partition the space would be kd-trees that splits the space parallel to one of the axis in each step. Those two sets of primitives are split recursively again until every leaf has only a maximum amount of primitives. While kd-trees offer slightly faster ray intersection tests than BVH, their disadvantages outweigh. BVHs can be built much faster and are generally more numerically robust, so it is less likely to happen to miss intersection duo to round-off errors than kd-trees are. Other desired traits of BVHs is that they require a maximum number of $2n - 1$ nodes for n primitives and, since we are splitting the primitives after a particular amount of primitives, we will never have the same primitive in two different nodes.

TODO: image?

2.6.1 BVH Construction

Obviously, the first step is the construction of the BVH data structure. The basic structure of BVH is a binary tree, with every node containing pointers to up to two children and every leaf holding a set maximum number of primitives. The construction technique that is most popular and also used in PBRT is the top-down construction. At the beginning of the construction we hold a set containing every single primitive of the scene. With each step, according to a split method we have chosen, we partition the primitives into two disjointed sets. Note that regardless of the split method, we chose, we will always split the primitives among a certain axis to retain locality. Examples for popular splitting heuristics are "Middle", which just partitions the room among the middle of a chosen axis, "EqualCounts", which partitions the primitives among a chosen axis so the disjointed sets have the same number of primitives, and SAH 2.6.1.1. While the first two split methods allow a easy and fast construction, their quality is very lacking. Recursively splitting a

nodes into two disjointed sets of primitives constructs the tree. When a certain node holds less than a defined number of primitives, we do not split again and instead call the node a leaf.

Every node needs to be holding onto some information in order to allow the intersection tests. Evidently, every inner node needs to have a reference to both of its children. As we have mentioned earlier, it also needs to store is a bounding volume that wraps around all its primitives. When testing, if we will need to intersect any of the primitives of a node with a ray, we will be instead, intersecting the box with the ray. Typically, we will be using minimum fit axis-aligned bounding boxes for this task.

2.6.1.1 Surface Area Heuristics

https://puu.sh/AU5HS/49a242e5f1.png

We have presented two split methods earlier in this paper. While they both do work well in some situations, they both have clear advantages, especially if the primitives are not evenly distributed over the scene. Looking at (TODO:ref), it is obvious, that both the "Middle" and "EqualCounts" split the scene in a way that can lead to very inefficient intersection tests. In the "Middle" split, although the triangle mesh on the right child node are focused on a very small space, the drawn ray will still make intersection test with each of the triangles of the right child node. Similarly, in the "EqualCounts" split, many unnecessary intersection tests are made. A desirable split would be the third split and we will now introduce a split method that favors these kinds of splits.

The surface area heuristics (SAH) defines a model that assigns a quality to a given split. The idea behind the SAH cost model is actually very simple. When we want to decide the best possible split for given primitives, first, we have to decide if it is perhaps better not to split at all. That means, when a ray traverses through the bounding box of the node, we would have to make an ray-primitive intersection test with each primitive. That means the cost c_g is

$$\sum_{i=1}^{N} t_i sect(i), \tag{2.18}$$

where N is the number of primitives and $t_i sect(i)$ the time required for the intersection test with the i-th primitive. For simplicity, we will assume, that every ray-primitive intersection test takes the same amount of time.

Clearly, we can also split the primitives. The cost $c(A, B)$ would be

$$c(A, B) = t_t rav + p_A \sum_{i=1}^{N_A} t_i sect(a_i) + p_B \sum_{i=1}^{N_B} t_i sect(b_i) \tag{2.19}$$

with t_trav being the additional overhead time to traverse through a child, p_A and p_B being the probabilities that the ray passes through the respective child node. The probabilities p_A and p_B can be calculated using their respective surface areas

$$p_C = \frac{s_G}{s_C}, \tag{2.20}$$

where s_G is the surface area of the node and s_C is the surface area of the child.

2.6.2 BVH Traversal

2.7 The algorithms for comparison

3 Own Data Structures

3.1 Node

3.2 Light Bounding Volume Hiarachy

4 Our Algorithm

4.1 General Idea

4.2 Tree Construction

4.3 Tree Traversal

5 Evaluation

6 Conclusion

…

Bibliography

[Jay03] E. T. Jaynes. *Probability Theory: The Logic of Science*. 1st ed. Cambridge University Press, 2003.

[Vea97] E. Veach. “Robust Monte Carlo Methods for Light Transport Simulation”. 1997.