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

Ray 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 computation s. 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 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 thesis we will introduce a light sampling technique that optimizes 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 ray 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. $p : D \to \mathbb{R}$ 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

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

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.

2.2 Monte Carlo Integration

When generating an image using ray 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.

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

$$F_N = \frac{b-a}{N} \sum_{i=1}^{N} f(X_i) \tag{2.5}$$

is equal to the integral. If we use a PDF $p(x)$ instead of an uniform distribution, the estimator

$$F_N = \frac{1}{N} \sum_{i=1}^{N} \frac{f(X_i)}{p(X_i)} \tag{2.6}$$

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.

While standard quadrature techniques converge faster than Monte Carlo integration, which converges at the rate of $O(\sqrt{N})$, it is the only integration method that allows us to deal with higher dimensions of the integrand. Later, we will explain why the light transport problem of the ray tracing algorithm is theoretically an infinite-dimensional problem. [Vea97]

2.3 Importance Sampling

2.4 Multiple Importance Sampling

2.5 Bounding Volume Hierarchies

2.6 Surface Area Heuristics

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.