Utilising path-vertex data to improve Monte Carlo global illumination.

Doidge, Ian Christopher

Utilising Path-Vertex Data to Improve Monte Carlo Global Illumination

Ian Christopher Doidge

Submitted to Swansea University in fulfilment
of the requirements for the Degree of Doctor of Philosophy

Swansea University
Prifysgol Abertawe

Department of Computer Science
Swansea University

2014

Declaration

This work has not been previously accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ... . (candidate)

Date 14th February 2014

Statement 1

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed .... (candidate)

Date 14th February 2014

Statement 2

I hereby give my consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed .... (candidate)

Date 14th February 2014

Abstract

Efficient techniques for photo-realistic rendering are in high demand across a wide array of industries. Notable applications include visual effects for film, entertainment and virtual reality. Less direct applications such as visualisation for architecture, lighting design and product development also rely on the synthesis of realistic and physically based illumination. Such applications assert ever increasing demands on light transport algorithms, requiring the computation of photo-realistic effects while handling complex geometry, light scattering models and illumination. Techniques based on Monte Carlo integration handle such scenarios elegantly and robustly, but despite seeing decades of focused research and wide commercial support, these methods and their derivatives still exhibit undesirable side effects that are yet to be resolved. In this thesis, Monte Carlo path tracing techniques are improved upon by utilizing path vertex data and intermediate radiance contributions readily available during rendering. This permits the development of novel progressive algorithms that render low noise global illumination while striving to maintain the desirable accuracy and convergence properties of unbiased methods.

The thesis starts by presenting a discussion into optical phenomenon, physically based rendering and achieving photo realistic image synthesis. This is followed by in-depth discussion of the published theoretical and practical research in this field, with a focus on stochastic methods and modern rendering methodologies. This provides insight into the issues surrounding Monte Carlo integration both in the general and rendering specific contexts, along with an appreciation for the complexities of solving global light transport. Alternative methods that aim to address these issues are discussed, providing an insight into modern rendering paradigms and their characteristics. Thus, an understanding of the key aspects is obtained, that is necessary to build up and discuss the novel research and contributions to the field developed throughout this thesis.

First, a path space filtering strategy is proposed that allows the path-based space of light transport to be classified into distinct subsets. This permits the novel combination of robust path tracing and recent progressive photon mapping algorithms to handle each subset based on the characteristics of the light transport in that space. This produces a hybrid progressive rendering technique that utilises the strengths of existing state of the art Monte Carlo and photon mapping methods to provide efficient and consistent rendering of complex scenes with vanishing bias.

The second original contribution is a probabilistic image-based filtering and sample clustering framework that provides high quality previews of global illumination whilst remaining aware of high frequency detail and features in geometry, materials and the incident illumina-

tion. As will be seen, the challenges of edge-aware noise reduction are numerous and long standing, particularly when identifying high frequency features in noisy illumination signals. Discontinuities such as hard shadows and glossy reflections are commonly overlooked by progressive filtering techniques, however by dividing path space into multiple layers, once again based on utilising path vertex data, the overlapping illumination of varying intensities, colours and frequencies is more effectively handled. Thus noise is removed from each layer independent of features present in the remaining path space, effectively preserving such features.

Acknowledgements

Firstly, my thanks must go to my supervisor, Dr. Mark Jones, for his guidance and wisdom throughout my studies. He has allowed me incredible freedom to pursue my own research directions, with seemingly resolute confidence. Were it not for his enthusiasm for computer graphics that I was exposed to in my undergraduate degree, I may not have developed my own passion for this most rewarding and frequently demanding of subjects. I would also like to extend thanks to my second supervisor Dr. Benjamin Mora for his guidance and suggestions towards my early research.

Thanks also go to my examiners Dr. Kurt Debattista and Dr. Gary Tam for their suggestions for improvements to the thesis. Thank you for making the examination process such a pleasant and enjoyable experience, and for reminding me that there is a plural of phenomenon.

My undergraduate degree included many hours in discussion with Prof. Min Chen, amongst others in the department at Swansea, from whom I have learnt a huge amount. I would also like to thank the many members of the department who made studying for a Ph.D. a far more enjoyable experience. In particular my thanks go to Matthew Edmunds, from whom I have learnt more than I ever need to know about CFD and flow visualisation. Matthew Gwynne for numerous debates on C++, alongside Phil James, Fredrik Forsberg and Liam O'Reilly for the much needed comedy, immaturity and theoretical ramblings; some of the key ingredients of student life.

Thanks would not be complete without mentioning my friends in and around Swansea, who made it such an amazing place to live. You will be sorely missed. To my family, both here and recently departed, whom I know are all immensely proud of me. Their ongoing pride and support have kept me going throughout my studies and given me the motivation to push myself ever further.

Lastly, I am indebted to my wife, Heather, for her continual support and endless patience over the years of my Ph.D., despite being a recurrent widow to paper submissions and thesis writing. Sharing in all the trials and tribulations of the last seven years of life, it is her unconditional support that has kept me going and for which I owe her a great deal. Although she will never read this thesis (or want to hear of it again), it is dedicated to her as a permanent reminder of the unwavering love, commitment and determination we have for each other.

Table of Contents

1 Introduction **1**
- 1.1 Demands for realistic rendering . . . 1
- 1.2 Objectives and research hypothesis . . . 3
- 1.3 Thesis organisation . . . 4
- 1.4 Achieving realism in rendering . . . 5
  - 1.4.1 Physically-based light transport . . . 5
  - 1.4.2 Models of light . . . 6
  - 1.4.3 Realism vs Photo-realism . . . 9
- 1.5 Radiometry and Photometry . . . 10
  - 1.5.1 Photometry . . . 11
- 1.6 Conclusion . . . 12

2 Physically-Based Rendering **13**
- 2.1 Ray casting and primitive intersections . . . 14
  - 2.1.1 Acceleration techniques . . . 14
  - 2.1.2 Specialised data structures . . . 17
- 2.2 Surface representation . . . 17
  - 2.2.1 Bi-directional Surface Distribution Functions . . . 18
  - 2.2.2 Texture mapping . . . 21
- 2.3 Light sources and emission . . . 21
- 2.4 Light transport . . . 21
  - 2.4.1 Ray tracing . . . 22
  - 2.4.2 Path space . . . 23
  - 2.4.3 Global illumination as an integration problem . . . 24

3 Monte Carlo Integration and Rendering **29**
- 3.1 Probability theory . . . 29
  - 3.1.1 Probability distributions and density functions . . . 30
  - 3.1.2 Expected values and variance . . . 31
- 3.2 Basic Monte Carlo integration . . . 31
- 3.3 Variance reduction techniques . . . 34
  - 3.3.1 Importance sampling . . . 34
  - 3.3.2 Generating continuous random variables . . . 36

3.3.3 Russian roulette and splitting . . . 37
3.3.4 Adaptive sample generation . . . 38
3.3.5 Markov-Chain Monte Carlo . . . 39
3.3.6 Additional techniques . . . 39

4 Paradigms for solving light transport **41**
4.1 Monte Carlo path tracing . . . 42
4.1.1 Building on MC ray tracing . . . 44
4.1.2 Path tracing . . . 45
4.1.3 Sampling and path densities . . . 49
4.1.4 Path termination . . . 52
4.1.5 Particle tracing . . . 52
4.1.6 Bi-Directional path tracing . . . 54
4.1.7 Metropolis Light Transport . . . 57
4.1.8 Trading accuracy for performance . . . 59
4.2 Photon Mapping . . . 60
4.2.1 Photon tracing and storage . . . 60
4.2.2 Radiance estimates . . . 62
4.2.3 Rendering . . . 63
4.2.4 Bias, variance and consistency . . . 64
4.2.5 Limitations of classical photon mapping . . . 66
4.2.6 Progressive Photon Mapping . . . 67
4.3 Alternative Monte Carlo methods . . . 72
4.4 Extending unbiased Monte Carlo methods . . . 73
4.4.1 Cache-based importance sampling . . . 74
4.4.2 Adaptive rendering . . . 75
4.4.3 Sample reconstruction . . . 76
4.4.4 Noise removal . . . 77
4.4.5 Image based filtering . . . 78
4.5 Conclusion . . . 81

5 Mixing Monte Carlo and Progressive Rendering **83**
5.1 Introduction . . . 83
5.2 Motivation . . . 85
5.2.1 Limitations of unbiased Monte Carlo methods . . . 86
5.2.2 Photon Mapping for caustic illumination . . . 88
5.2.3 Hybrid approaches . . . 90
5.3 Overview . . . 92
5.4 Path space separation and filtering . . . 93
5.5 Path tracing . . . 94
5.6 Caustic evaluation . . . 97
5.6.1 Photon tracing . . . 98
5.6.2 Photon gathering . . . 100
5.7 Results and discussion . . . 101

5.7.1 Cornell Box . . . 103
5.7.2 Metallic Ring - difficult caustics . . . 104
5.7.3 Shapes - Reflective open environment . . . 105
5.7.4 Quantitative comparisons . . . 108
5.8 Limitations and future work . . . 110
5.9 Conclusions . . . 114

6 Irradiance-Aware filtering for Monte Carlo rendering 119
6.1 Introduction . . . 119
6.1.1 Irradiance awareness in image based filtering . . . 120
6.1.2 Contributions . . . 122
6.1.3 Poisson Distribution . . . 123
6.2 Rendering and sample clustering . . . 124
6.2.1 Cluster formation . . . 127
6.3 Noise removal and filtering . . . 132
6.3.1 High intensity sample rejection . . . 133
6.3.2 Illumination preserving filter . . . 139
6.3.3 Pixel radiance computation . . . 146
6.4 Results and discussion . . . 149
6.5 Limitations and future work . . . 156
6.6 Conclusions . . . 158

7 Conclusion 161

Chapter 1

Introduction

Contents

1.1 Demands for realistic rendering 1
1.2 Objectives and research hypothesis 3
1.3 Thesis organisation 4
1.4 Achieving realism in rendering 5
1.4.1 Physically-based light transport 5
1.4.2 Models of light 6
1.4.3 Realism vs Photo-realism 9
1.5 Radiometry and Photometry 10
1.5.1 Photometry 11
1.6 Conclusion 12

1.1 Demands for realistic rendering

Efficient techniques for photo-realistic rendering are in high demand across a wide array of industries. As global illumination research has matured, these demands and the areas of application for such methods have increased dramatically. Computer generated imagery and visual effects for the film industry are commonplace [MHH+12], and have reached such a level that they can go unnoticed when composited with real life footage; the ultimate aim of photo-realistic rendering. Additionally, the expectations for accurate light transport in animated films are ever increasing, restricted by the computational requirements to provide practical frame rates for production, and far less so by their accuracy. Global illumination is now a common and viable option for secondary markets such as advertising, product development and design, where image synthesis is a tool rather than an end product. Computer Aided Design (CAD) used in architecture, lighting design and product development requires highly accurate light transport simulation and material models to allow artists and designers to accelerate the time from concept to final product, which can remove the need to conduct tests in the physical world, and useful for accelerated marketing and focus groups. Physically-based simulations in virtual

environments can be relied upon instead to provide accurate visual and numerical data without the costs and time involved in creation and procurement of the physical resources. Global illumination is finding increased interest from the data visualisation community due to recent performance improvements, especially for scientific and medical visualisation [GBP08, JKRY12]. Global lighting models provide improvements in depth perception and spatial-acuity, and illumination for volumetric data in the form of subsurface scattering, exceeding what is possible with conventional surface shaders and aligning well with the widespread use of volume ray casting [BB07]. Even in the realms of real-time, frame rate critical applications such as video games, global illumination is looking to be a feasible alternative to the rasterisation pipeline used extensively today [Bik13].

These numerous applications assert ever increasing demands on rendering algorithms, requiring photo-realistic image synthesis from environments containing complex geometry and material models, incorporating a wide range of scattering phenomenon and illumination features. Additional pressures come from industry standards such as increased display resolutions, reaching pixel counts in excess of 8 million [Dig12]. Improvements in modern hardware and the accessibility of massively parallel computing through GPGPU and OpenCL [Khr13] have opened up new possibilities for global illumination to both individuals and for general use outside visual computing centric industries.

Despite its widespread use, performance remains the limiting factor for global illumination in the majority of applications. Highly accurate methods require significant computational effort, in some cases requiring hours (or days) to produce error and noise free results. The major bottleneck lies in computing the light transport between surfaces in the virtual environment. Just as in the physical world, light is distributed around the scene through emission, scattering and absorption until detection occurs; for example by a camera sensor. However light arriving at a surface location not only comes directly from light sources, but also from light scattered from other surfaces visible from that location. In turn, the light arriving from those surfaces may also have arrived via scattering, and so forth. The chaotic and incoherent distribution of light makes such simulation difficult, and heightens the complexity of light transport algorithms. Measuring the light energy that reaches detectors in the scene (our virtual camera) whilst striving for realistic image synthesis comes at high cost, as will become apparent throughout this thesis.

To avoid the resultant computation time, approximations are often developed to satisfy the trade off between rendering quality and attaining acceptable frame rates. The primary focus in light transport research has thus been to reduce the necessity of this trade off and has seen, and continues to see, significant attention from researchers. By relaxing the constraints placed upon light transport algorithms, assumptions can be made to provide approximations to the true solution of the light transport simulation. Broadly speaking, accomplishing this can follow two approaches in which algorithms can be:

- Developed with a focus on performance and aim to achieve the highest possible accuracy within a required frame rate, or;
- Build upon robust and highly accurate methods, and improve on efficiency using techniques that strive to reduce visible error.

Rasterisation is an extreme example of this first approach, having been used as the industry standard for interactive and real-time rendering for decades. At its most basic level, rasterisation does not account for light source occlusion, hence demanding applications such as modern game engines must rely on a multitude of extensions to provide significant improvements in quality [Cry13]. The reliance on rasterisation adds to the complexity of implementation and development of algorithms to approximate realistic phenomenon efficiently, as the rendering pipeline becomes composed of many complex parts, each requiring attention during the development cycle. Using a ray tracing approach to physically-based global illumination provides a natural alternative that can elegantly and robustly handle complex effects under a single model, whilst alleviating many restrictions that are placed on content creators by non-realistic materials and incomplete light transport models. Inaccurate and restrictive algorithms will therefore only become more robust by extension, as permissible by the performance increase provided by hardware developments.

When considering the second approach, highly accurate methods exist that are physically based (discussed following this introduction) provide superior quality and extensibility. Additionally, a choice can be made to trade a measure of accuracy or robustness in such methods to reduce noise during the early stages of rendering, whilst maintaining an underlying error free solution in the limit. Given trends in hardware performance and increasing parallelism over the last decade, accurate global illumination methods will inevitably become feasible even without additional research effort in the field of light transport. It is this natural evolution of hardware architecture that has been largely responsible for the popularity of ray tracing approaches thus far, which for which interactive frame rates have been achievable on commodity hardware, even without the research of the last decade [WS01]. Coverage of such techniques will be provided since they are the underlying motivation behind work in this thesis.

1.2 Objectives and research hypothesis

One side effect of global light transport's complexity is that a significant quantity of intermediate data is calculated pertaining to local light transport, as the scattering of light between surfaces is calculated. However, much of this data is computed incoherently throughout the rendering process. Since it is typically not of interest after contributing to the value of an image pixel, such data is discarded despite being both large in quantity and accurate with respect to the local and global light transport of the simulation. Whilst contributing only in small part to the appearance of the final image, together it may be possible to utilise such data to evaluate light transport more efficiently, or provide a stronger indication of the illumination over a local region. Thus it is the aim of this thesis to:

- Develop novel techniques to both identify and make use of the intermediate data available during global light transport algorithms,

- Using this data, reduce the visible noise present in rendered images to improve the efficiency of global light transport algorithms.

In addition, the relevance and impact of the contributions of this thesis need to be considered with respect to the field of research. As discussed, building upon accurate and robust algorithms is a highly future orientated paradigm, thus further objectives of this thesis are to:

- Build upon on Monte Carlo path tracing algorithms that are already capable of providing unbiased solutions to the light transport problem.
- Strive for the development of algorithms that fit into the paradigm of progressive rendering, in order to maximise their use and potential impact.

Currently, Monte Carlo methods provide among the strongest approaches for handling realistic models of geometry, light scattering and visual phenomenon. Alongside their relative simplicity, it is this high quality and generality that makes them an elegant solution for global illumination and hence their wide adoption in rendering systems and for commercial applications [Lux13, Jak13, NVI13] thus far. Following the objectives and hypotheses above, this thesis develops techniques to address some of the outstanding issues facing such approaches to the light transport problem which will become apparent in the following chapters.

By achieving such objectives, the resulting thesis will have considered the longevity of any proposed techniques with respect to future research directions and current trends, going some way to ensure the impact of this work in the field is maximised.

1.3 Thesis organisation

The remainder of this chapter introduces and motivates the use of physically based rendering for photorealistic rendering and some insight into the complexities of light transport for simulating optical phenomenon.

Chapter 2 provides a short introduction to the concepts of a ray-based light transport framework, along with its benefits and drawbacks from a practical perspective. State-of-the-art techniques for accelerating ray tracing are discussed with direction to recent literature, providing an appreciation of the performance bottlenecks.

Monte Carlo integration and its properties are discussed in Chapter 3 from a theoretical perspective, which forms the backbone of the modern rendering techniques built upon in this thesis. The remainder of the chapter discusses existing variance reduction techniques adopted for reducing noise during rendering.

An insight into the range of state-of-the-art methodologies and paradigms for global illumination light transport is presented in Chapter 4 based upon the theory outlined in the previous chapters. Implementation and practical details of the well known family of unbiased Monte Carlo path tracing algorithms are provided, upon which the contributions in this thesis are built, along with some of the more successful extensions that have been proposed to date. The remainder of Chapter 4 provides detail of state-of-the-art photon mapping techniques, along with alternative popular paradigms and the role and effects of introducing bias in rendering. This will provide a sound context and motivation from which the novel work proposed in the remaining chapters can be discussed.

Chapter 5 presents the initial novel contributions to the field, improving upon Monte Carlo path tracing methods. This new algorithm proposes a hybrid strategy, partitioning the complex space of light transport during rendering, in order to leverage the strengths of unbiased path tracing and progressive density estimation methods which individually experience slow convergence and high noise levels under complex lighting conditions. To motivate this work, the issues plaguing the state-of-the-art methods in these two paradigms are discussed. Existing literature related to path space separation and hybrid approaches is discussed with reference to the new work. Details of the algorithm are then delved into, from theoretical and practical angles, providing a thorough understanding into its operation and decisions for design and implementation. The chapter is concluded by providing visual and numerical comparisons to related rendering algorithms showing its usefulness and applicability to global illumination.

The second set of novel contributions form the basis for Chapter 6. An irradiance-based filtering framework is proposed, designed around the need to detect image discontinuities and edges in both high and low variance sources. This is crucial for reducing bias and artefacts in the image when applying noise reduction filters. First, the difficulties in designing image-based filtering methods is presented using examples from existing work, and the significance of such techniques in realistic rendering is discussed. The main content of the chapter introduces a new clustering framework, allowing the compact storage of illumination contributions whilst enabling visual discontinuities to be detected. A novel filtering algorithm makes use of such clusters to preserve complex surface detail and respects illumination edges via a novel probabilistic approach. A number of visual comparisons are made demonstrating the usefulness of the technique for complex materials and textures.

This thesis concludes with a summary of the techniques developed in chapters 5 and 6, a discussion of their impact and potential for future work along with some insight into the still open and more general problems prevalent in photo-realistic rendering.

1.4 Achieving realism in rendering

In order to improve the quality of light transport simulation and strive for realism, it is important to determine what qualities are essential for tricking the human visual system into believing the reality of synthesised images, and those that can easily destroy such perceptions. In this section, types of visual phenomenon and cues are discussed that commonly present themselves, along with those that are less frequently observed or that our visual systems are less sensitive to. This provides an abstract model of light transport that both maximises its potential for realism and minimises complexity, making implementation easier.

1.4.1 Physically-based light transport

Techniques in rendering are focussed towards achieving photo-realism (within some constraints), but the way in which this is accomplished can vary greatly. Thus to understand the place of this thesis in the vast field of rendering, an important distinction must be made. Many algorithms such as ambient occlusion techniques [MFS09] attempt to produce photo-realistic effects but through the use of arbitrary illumination terms and unit-less quantities. Although they can be convincing under certain conditions, this lack of realistic foundations mean they cannot easily

achieve photo-realism, requiring considerable parameter tweaking to obtain the desired result [MHH+12].

Conversely, physically based rendering techniques have more rigorous theoretical and mathematical underpinnings. They can be thought of as numerical simulations of the behaviour of light within a virtual environment, designed to be as close to physically valid as possible. Physical laws such as the conservation of energy and optics in the real world thus form the crux of such techniques, partnered with units of measurement from radiometry. Consequently, they can provide unrivalled realism and an intuitive method of achieving realistic results.

1.4.2 Models of light

The key to efficient synthesis of photo-realistic imagery in practice is tightly coupled to the concept of perceptual necessity. Whilst science has a seemingly rigorous understanding of optical physics and electromagnetic radiation down to the quantum level, and could endeavour to simulate it accurately, it is unnecessary for the majority of applications. Instead, an abstraction can be created, simulating only the phenomenon that are evident in the output for a particular application. In this work and in the majority of computer graphics literature, this abstraction is based on the physical effects that manifest themselves as the visible phenomenon observed in everyday life. Hence, it does not fit neatly into a standard optical physics model (itself comprised of both wave and particle theories), so must be drawn from multiple models. Here, a brief overview of the noticeable lighting phenomenon observed around real environments, and a discussion on their typical relevance and development in physically based rendering is provided.

1.4.2.1 Geometrical Optics

The majority of optical effects observed in everyday environments, and hence simulated in computer graphics, can be explained using a geometrical (or ray based) model. Such a model follows the particle like behaviour of light, where it is assumed to travel in straight lines instantaneously between points. This geometric model provides an excellent simplification of light propagation when dealing with objects and features significantly larger than the wavelength of visible light, commonly the case in rendering for scenes of macroscopic scale.

Reflection is the primary mechanism that allows visual systems to obtain information about their surroundings, and can be fundamentally categorised into *specular* and *diffuse* scattering. Specular (mirror-like) reflections follow the law of reflection in which for any incoming ray of light there is only one outgoing direction. Despite its simplicity it is not trivial in a rendering context, since obtaining clear mirror like reflections requires accurate evaluation of the lighting along the outgoing direction, requiring a multi-pass or recursive approach. Diffuse reflection occurs, when light arriving at a surface is scattered in many directions on account of numerous small scale interactions, some of which is reflected towards the eye. Due to their dominance in common environments, diffuse reflections are the basis of most shading algorithms in computer graphics, providing a sense of depth and making geometric detail visible. In physically-based rendering, the appearance of a wide range of materials can be modelled by distribution func-

tions which represent the proportion of light reflected in a particular direction. Such functions vary greatly in complexity, and commonly used distributions are covered Chapter 2.

Participating Media accounts for the scattering, absorption and emission of light in 3D space between surfaces. Typical examples are waterborne or airborne particles such as dust, fog and smoke. Participating media is responsible for effects in atmospheric optics such as sky appearance and colour shifts seen at sunrise and sunset, as well as rarer events such as crepuscular rays [TV08, LL01]. In computer graphics, realistic analytical daylight models [PSS99, HMS05, HW12], have enabled accurate rendering under natural lighting conditions and phenomenon such as rainbows and halos have also seen physically-based solutions [REK+04, KFY+10]. Light transport for participating media is classified into two main groups; homogeneous or inhomogeneous. The scattering, absorption and emission of homogeneous volumes remain constant throughout their spatial-extent, allowing them to be modelled in closed-form. Inhomogeneous media are more complex, normally requiring the evaluation of integrals via Monte Carlo methods [YIC+10]. Often, participating media is absent or ignored since its effect on the image is not offset by the additional computation required to simulate it. Commonly, extensions to incorporate homogeneous participating media are relatively simple in ray based light transport methods, since the effect of absorption and scattering can be computed along a ray with known endpoints.

Subsurface scattering is closely related to participating media, but describes the scattering of light under the surface of translucent objects, modelled as an extension to reflection models [JMLH01]. Light penetrates the surface of translucent and transparent materials undergoing scattering and absorption internally, exiting at a different point on the objects surface. This enables materials such as porcelain and marble to be modelled more realistically and the visual appearance of many organic materials such as skin is also dominated by subsurface scattering [Tuc07] making it an important aspect of realistic rendering.

Refraction is the bending of light at surface boundaries when light passes between medium. Due to the laws of energy conservation, its phase velocity changes according to the density change between the two medium and the incident angle at the boundary. Systems of lenses in cameras, microscopes and the human eye rely on refraction to manipulate light for magnifying and focussing images on the sensor or retina. The behaviour of light at such interfaces can be modelled using the Fresnel equations [Gla95], predicting the fraction of light that should be reflected or refracted at the boundary for a light ray incident at a given angle.

Refraction is also responsible for mirages and heat haze commonly seen above hot sands or tarmac caused by the refractive index of air changing due to variation in the air temperature, as heat is convected away from the hot surface. This refraction in so called non-constant media do not occur at surface boundaries and are inhomogeneous due to the variations in temperature, breaking down the assumption that light travels in straight lines. This makes visibility queries challenging and due to their rarity, such complex effects are generally ignored.

Dispersion is the result of visible light becoming spatially separate into its component colours when undergoing refraction through so called dispersive media. This is a result of the wavelength dependent response to interactions with certain materials. Rainbows are a naturally occurring example, as well as the brilliance associated with cut diamond, but such effects also play an important role in industries such as lens design and fibre optic communications, leading to artefacts such as chromatic aberration. Rendering algorithms commonly treat light as monochromatic, assuming light behaves as a single wavelength in order to simplify computation. Ray based rendering can account for dispersion on demand, creating one or more monochromatic rays to replace a single polychromatic ray when dispersion occurs, discussion of which can be found in [SFD00].

1.4.2.2 Light as Waves

Light is also known to have wave like behaviours, approximated by Maxwell's equations of electromagnetism [Gla95] that account for more complex phenomenon generally explained by the presence of electric and magnetic fields. Less noticeable phenomenon such as diffraction, interference and polarisation can be explained under this model. Though commonly not accounted for in mainstream rendering, such effects are still important from a physical realism standpoint and require consideration during material and implementation design.

Diffraction is the bending of light around near field obstacles. The famous double slit experiment of Thomas Young was an example of such a phenomena, and strongly supported the theory that light consisted of waves. Diffraction generally only leads to observable phenomena when the relative size of the objects is close to the wavelength of visible light, far smaller than the scale usually dealt with in computer graphics. Despite this examples such as the colourful reflections from a compact disc and the blooming effect from strongly backlit objects have seen increased research attention in recent years. Stam [Sta99] developed a shader model for local surface based diffraction. Recently, Zhang and Levoy [ZL09] linked techniques used in geometric optical simulations to light field representations used in computer graphics allowing Oh et al.[OKG+10] and Cuypers et al.[CHB+12] to formulate local reflectance functions that can be used to implicitly simulate global interference patterns through the addition of phase information to each ray. This allows for far simpler integration of diffraction effects into existing ray-based global illumination renderers without modification, making use of existing sampling techniques for acceleration.

Phase and Coherence The effects of diffraction are observable due to wave coherence, a product of correlation between the phases of light waves in which their effect is either increased or decreased non-linearly depending on their phase difference. The majority of light is incoherent, and as such the intensity of light rays along a given trajectory scales linearly with the number of rays. For coherent rays this is not the case, and leads to interference patterns such as those observed during Young's double slit experiment. Such effects are important in rendering thin transparent films, such as oil spills or anti-glare coatings on optical lenses, where a light ray can be internally reflected several times within a space similar to its wave-

length, superimposing upon itself before escaping the film. The synthesis of such effects is approximated in some modern rendering frameworks such as LuxRender [Lux13].

Polarisation Polarisation is arguably the most relevant phenomenon for computer graphics that arises from a wave based model, and can be thought of as the superposition of two transverse waves with the same trajectory, oscillating perpendicular to one another. Unpolarised light exhibits little correlation between the phase of these two waves, and upon scattering these waves may undergo changes in phase as a result of the material's electric field, causing them to become polarised to some degree. Photographers and sunglasses manufacturers use polarising filters to reduce light transmission from reflections and the sky, blocking waves of certain polarisations, increasing contrast and colour saturation whilst reducing glare. Fresnel equations that model refraction and reflection are affected by polarisation, and it can play an important role in accurately rendering inter-reflections. In depth discussions and implementations of polarisation based ray-tracers have been presented recently in SIGGRAPH courses by Wilkie et al.[WW11, WW12].

1.4.2.3 Quantum Effects

At its most fundamental level, light can be thought of as discrete packets of energy, or quanta, that can describe fundamental properties of energy across many areas of physics. In light transport, there are few visible effects that require such fundamental principles. The most well known and commonly observed of these are fluorescence and phosphorescence; photo luminescent effects whereby light of a given wavelength is absorbed and emitted at a different visible wavelength. However, since time dependent effects are rarely seen and the steady-state model used in rendering is highly convenient for all other light transport, they are often ignored. Effects arising from magnetic or gravitational fields, non-Newtonian physics and the theories of relativity are ignored, since they are not applicable to general applications, although some have seen ray tracing based simulation [TM10].

1.4.3 Realism vs Photo-realism

The goal of photo-realism is to produce imagery of a virtual environment that is indistinguishable from a photograph of the same scene in the real world. The need for accurate models of light transport and materials has been discussed, that simulate phenomena observable by the human visual system. Natural lighting models in the allow us to reproduce realistic conditions, and man made light sources can be modelled via accurate geometric representation and manufacturers lighting data. Lack of realism can also stem from artistic elements such as absence of dust, dirt and weathering in a virtual environment, or overly perfect geometry that is free of the deformations either found naturally or from imperfect manufacturing processes. Although true realism of the human visual system may be the ultimate goal, this is fundamentally difficult to both achieve and verify. The organic and interpretive nature of the human sensory system dictates that no quantitative comparisons can be made between synthesised imagery and real world environments. Thus even if the aim is to obtain techniques that produce truly realistic imagery, how would such techniques be evaluated? Electing instead to attain photo-realism means that

Quantity	Name	SI Units	Photometric Equivalent
Q	Radiant Energy	J	Luminous energy
Φ	Flux (Radiant power)	W	Luminous Flux (Luminous Power)
E	Irradiance	Wm^{-2}	Illuminance
M	Radiant exitance (or Radiosity)	Wm^{-2}	Luminosity
L	Radiance	$W \cdot sr^{-1} \cdot m^{-2}$	Luminance

Table 1.1: *Radiometric terms and their photometric equivalents used throughout this thesis, along with their symbols and SI units.*

results can be validated against quantitative data from actual camera sensors, providing like for like comparisons.

If the desire of photo-realistic rendering is to recreate scenes as if obtained through photography, then it is also necessary to simulate the characteristics of camera and lens systems that record such images. Camera flaws and imperfections come under the realm of rendering rather than artistic realism, since they are derived from physical effects of light transport through the lenses and apertures. Artefacts such as chromatic aberration, vignetting, bloom and glare come into this category, but as yet there has been limited research into the physically-based simulation of such effects in conjunction with global illumination or ray-based techniques. Lee et al.[LES10] explore real-time ray tracing based methods to emulate multiple lens effects including aberration and focussing effects such as curvature of field and tilt-shift. Hullin et al.[HESL11] present a physically-based lens-flare model, rendering convincing and complex lens flare effects with reference to existing camera lenses. Being inherently free from lens artefacts and having the capability to synthesise such imperfections on demand enables photo-realistic rendering to produce results rivalling those obtained with expensive camera and lighting equipment, and for many years has provided an increasingly cost effective alternative to conventional photography.

1.5 Radiometry and Photometry

In solving global illumination, the steady-state distribution of light within an environment is computed, assuming instant propagation between surfaces. Synthesising the phenomenon discussed above requires an understanding of the physical quantities that describe how light energy is distributed. The field of radiometry is concerned with the measurement of electromagnetic radiation, and an overview of relevant theory is discussed in this section. An overview of the main radiometric quantities are outlined in Table 1.1. Flux Φ, measured in Watts ($W = Js^{-1}$), is the basic unit of energy in rendering since a time independent model is assumed. To compute light's interaction with objects, the flux flowing through a surface per unit area, known as *flux density* needs to be computed. It is often useful to differentiate between flux arriving and leaving a surface, referred to as *irradiance* and *radiance exitance* respectively.

Flux density is expressed in Wm^{-2}:

$$\frac{d\Phi}{dA}$$

Flux can also be expressed with respect to *solid angle*; the 2D area subtended by a region from a point x in space, projected on the unit sphere at that point. Using solid angle as opposed to unit area is convenient when dealing with point-based light transport and abstractions such as point light sources for which surface area is not well defined or differentiable. Perhaps the most fundamental and most useful quantities in rendering is *radiance*. Radiance is the flux density per unit projected area, per solid angle:

$$L = \frac{d\Phi}{d\omega dA^{\perp}} = \frac{d\Phi}{d\omega dA \cos\theta}$$

The area measure dA is projected perpendicular to the direction ω that we are concerned with to form $dA^{\perp}$. However we also need to account for the angle of ω with the surface normal at the point x, since the flux arriving through $dA^{\perp}$ is spread out over a larger area than $dA^{\perp}$. The cosine term handles this change in area.

Using radiance allows us to simplify the modelling of light transport. Expressing quantities in radiance allows the computation of point to point energy transfer, invariant along straight paths and which does not decrease with distance. This makes it a natural quantity to use in partnership with ray tracing for solving visibility and occlusion queries. Furthermore, given the radiance at a point, all other measurements can be calculated as an integral of radiance over direction or area. It is therefore a useful quantity for the basis of a single rendering framework designed to handle many types of light transport algorithm.

1.5.1 Photometry

For most applications of realistic rendering discussed in this thesis, the requirements for realism are based upon the ability of the human visual system to detect non-realistic effects and artefacts. It is therefore logical to touch on the implications of visual perception and how radiometric quantities are perceived by the human visual system.

Photometry is the perceptual measurement of radiometric quantities with respect to the human visual system. Due to the receptors that compose the human eye, its sensitivity is wavelength dependent, this differences between radiometric measurements do not reflect the perceptual disparities humans observe. Photometric models attempt to weight the measured radiance at each wavelength to represent the eye's non-linear response in a more linear fashion.

Each radiometric quantity has a photometric equivalent. Accurate conversion of quantities between the two systems requires integration over spectral response curves, however due to the properties of the receptor types that make up the human eye, this response curve is highly dependent on environment and context. Often these models are too costly to apply to rendering, and so approximations are used, ignoring contextual influences. A convenient representation for rendering are perceptual colour spaces, such as CIE XYZ and its derivatives, whose tristimulus values are designed to be analogous the trichromatic vision of the human eye. Conversion between colour spaces such as RGB to CIE XYZ are simpler and more efficient for computing perceptual differences during rendering.

1.5.1.1 Visual Perception

Quantitative measurements in photometry are useful, but to leverage the limitations of the human visual system in rendering requires complex models derived from experimental psychophysical data. How humans interpret changes in illumination depends on many spatial factors (and temporal factors in animation), including the brightness of the environment, local contrast and dynamic range. More discussion of perceptual phenomenon and models used in computer graphics can be found in [FPSG96, MTAS01, CCM03, RPG99]. Complex models begin to account for phenomenon such as threshold and contrast sensitivity, visual attention and temporal effects, providing adaptive rendering methods that concentrate on perceptual artefacts, as opposed to regions exhibiting potentially imperceptible numerical error. Such techniques have shown considerable efficiency improvements, and further investigation into easy to integrate spatial and temporal perception models coupled with smart adaptive rendering may provide more widespread benefits to rendering.

1.6 Conclusion

This chapter has introduced important concepts for understanding light transport phenomenon, their origins in optical physics theory, and the complexity of simulating them accurately in virtual environments. Radiometric quantities relevant to global illumination and the work presented in this thesis have been presented, and the importance of their photometric counterparts has been discussed. This chapter also outlined the importance of visual perception in rendering for many applications, both as an indication of realism and a tool for improving performance for global illumination. It is these realism and efficiency characteristics that dictate the success and widespread adoption of ray based global illumination in numerous industries and for non-commercial purposes.

Chapter 2

Physically-Based Rendering

Contents

2.1 Ray casting and primitive intersections **14**
2.1.1 Acceleration techniques 14
2.1.2 Specialised data structures 17
2.2 Surface representation **17**
2.2.1 Bi-directional Surface Distribution Functions 18
2.2.2 Texture mapping 21
2.3 Light sources and emission **21**
2.4 Light transport **21**
2.4.1 Ray tracing 22
2.4.2 Path space 23
2.4.3 Global illumination as an integration problem 24

Although the use of Monte Carlo methods to evaluate light transport forms the backbone of this work, there are a great many factors that influence the perceived realism of a scene in any rendering system and its overall performance. The representation of each component of a virtual environment must be considered in order to simulate its light transport effectively and efficiently. The design and modelling of geometry, light sources and the light scattering properties of materials in addition to the virtual camera model all have significant influence on the perceived realism and visual appeal of the final image. While highly accurate models of each are desirable, they can be impractical and the trade-offs between realism and performance need to be considered.

In this chapter the most significant of these components will be discussed, and a brief overview of their mathematical representations will be presented to provide a background to physically based rendering systems. Reference to recent relevant literature and texts for further reading will be provided.

2.1 Ray casting and primitive intersections

Building upon the largely ray-based model of light established in Section 1.4, the most fundamental definition is that of a ray segment representing the path of a light ray in the scene. Parametrically a ray segment can be represented by:

$$r(o,\mathbf{d}) = o + t \cdot \mathbf{d}$$

where o is the ray's origin, d is its direction in 3D space and $t \in [0,\infty)^{\mathbb{R}}$ is the distance along the ray which, with d, dictates the end point of the ray segment.

The initial challenge is to identify for a given ray which object in the virtual environment (if any) the ray intersects first from o to determine visibility (known as *ray casting*). Perhaps equally importantly, the precise point on the object the ray hits and the distance between the ray's origin and the object intersection (t) need to be established in order to propagate light around the scene and accurately compute shading. As a result the ray casting function can be defined as:

$$r(o,\mathbf{d}) = x \in S | x = o + \min\{t | (o + t \cdot \mathbf{d}) \in S\} \cdot \mathbf{d}$$

such that x is the closest ray intersection point on the scene geometry S from o. This is fundamental to discussing the simulation of light transport throughout the thesis.

Naïvely, evaluating this function requires the ray to be tested with every object in the scene. For anything but non-trivial scenes this is impractical due to the computational cost of ray-geometry intersections, which is typically the most expensive aspect of any ray tracing based renderer. Typically the term ray tracing is used to describe algorithms based on ray casting principles, which trace rays recursively or account for some level of secondary effects (in our case illumination) in addition to visibility.

Modern scenes can typically contain in the order of 10^4 to 10^6 individual geometric primitives, G, (shapes such as triangles, quadrilaterals or parametrics), and G can reach 10^8 for highly detailed models. To produce a basic ray cast image of N pixels, assuming constant illumination requires a minimum of one ray to evaluate Equation 2.1 for each pixel. For high resolution renders, N is often in the order of 10^7 pixels, and the need for additional rays R, to account for anti-aliasing, light source occlusion tests and global illumination can easily increase the total number of rays by an order of magnitude above the N pixel count of the image. Thus, effective acceleration techniques are critical to the success of ray tracing to reduce its naïve complexity of $O(N \cdot G \cdot R)$.

2.1.1 Acceleration techniques

As the earliest set back for ray tracing's widespread use, the need to accelerate and reducing the computation required for visibility testing is still a challenge tackled by many researchers today. The most effective techniques to date employ spatial partitioning, to subdivide the scene space into disjoint subsets, grouping primitives within a given volume. If no ray intersection is found for the bounds of a given partition, the primitives within that set do not need to be checked further, allowing rapid rejection. Such schemes incur small computational overheads due to testing against the bounding volume, but the overall computational savings are

far greater if the bounding volumes are constructed carefully. Havran [Hav00] notes that the quality of such partitioning schemes is reliant on two key properties:

- Relative volume: If a ray intersects the bounding volume, then the probability that the ray also intersects the object should be maximised.
- Efficiency: The intersection test between a ray and the bounding volume should be as computationally cheap as possible.

Havran's Ph.D. thesis [Hav00] provides a thorough discussion on ray casting and acceleration data structures with particular insight into traversal algorithms.

Such criteria have significant implications in the design and implementation of acceleration data structures, especially when dealing with hierarchical approaches which provide further improvements. The remainder of this section will give an overview of the predominant techniques in used today and provide reference to the literature for further reading.

2.1.1.1 kD-Trees

The k-Dimensional tree *kD-tree* belongs to the family of binary spatial partitioning trees (BSP-trees) and was introduced by Bentley [Ben75, FBF77]. It is widely used in search problems including clustering, density estimation and image processing [AGDL09] due to a number of desirable properties including:

- application to arbitrary dimensions (hence kD)
- providing adaptive partitioning of the domain to suit arbitrary data distributions
- supporting a flexible, heuristic-based approach to construction
- achieving fast traversal due to axis-aligned split planes

It consists of a binary tree in which each non-leaf node contained an axis aligned hyperplane spatially dividing the domain and hence the data elements within that node. They are constructed in a top down fashion such that the root node of the tree encompasses the entire domain. Construction of the tree is performed through a series of repeated steps:

1. A dimension $d \in k$ and data element e in the current node are chosen according to a cost heuristic to determine the location of the partition.
2. The data elements in the node are classified into two sets according to their location in d with respect to the split plane.
3. Bounds are computed for the two child nodes based on the split plane, and the data elements passed to the correct child.
4. Termination criteria are checked and if not met, the process is repeated from step 1 with the two child nodes.

The kD-tree has been one of the primary data structures in ray tracing since its introduction, and the cost heuristic is arguably the most influential component of the kD-trees success

and hence has seen much research [FS88, SF90b, SF90a]. MacDonald and Booth [MB90] introduce the surface area heuristic (SAH), that is a locally greedy cost model that attempts to maximise the ratio of data elements to node surface area, in line with the properties discussed above. A special case of such a heuristic is that empty regions of the domain can be cut off more favourably, making traversal of these regions almost free [Hav00, HKRS02].

More recently, focus has shifted towards improved construction times for acceleration data structures, with the accessibility of ray tracing dynamic scenes at interactive and real-time frame rates, which are dependent on the total time from the geometry updates through to final image rendering, incorporating the construction times as opposed to just query performance. Thus research focussed on achieving asymptotically faster kD-tree construction in $O(N \log N)$ time [WH06] and using approximation techniques to speed up the classification step [PGSS06, HMS06] and to provide parallel construction techniques for kD-trees [ZHWG08, SSK07]. This shift towards deformable and fully dynamic scenes has popularised alternative schemes for partitioning schemes as a replacement to the kD-tree.

2.1.1.2 Bounding Volume Hierarchies

Bounding volume hierarchies (BVHs), introduced to ray tracing by Goldsmith and Salmon [GS87], share many properties with kD-trees such as their adaptive nature and reliance on cost based heuristics for construction. However while kD-trees rely on spatial-division, and produce spatially disjoint sets, BVHs work via object-division thus child nodes are disjoint with respect to the geometry they contain, but can overlap in the spatial domain. Spatial subdivision like the kD-tree can result in duplicating references to geometric data held in the tree if it overlaps a given split plane, increasing memory consumption relative to object subdivision approaches like the BVH. As well as improved memory usage, which is often the bottle neck for modern hardware, object-subdivision permits additional flexibility when dealing with deformable geometry where small changes in geometry can be handled by incremental updates rather than a full rebuild of the data structure [Gar08]. Conversely, since they overlap spatially the traversal costs and hence query performance for BVHs can be reduced, which can be exacerbated under non-uniformly tessellated geometry. As a result a number of techniques have addressed this using a subdivision pre-process over the geometry [EG07, DK08] but this adds computation and memory usage that can be difficult to manage. Like kD-trees, the construction of BVHs has also seen attention [WBS07, WIP08].

Characteristics and advantages of both object-partitioning and spatial-partitioning have been researched and debated for many years, but more recently hybrid approaches have been proposed in an attempt to minimise the drawbacks of each. Stich et al.[SFD09] introduce spatial splits to BVH construction, evaluating both BVH and kD-tree style splits for each node, in a single data structure. This can provide improvements for static data or computing global illumination, where the construction cost of the data structure is a small proportion of the overall time to image. For dynamic scenes and interactive ray tracing, the near $O(n \log n)$ construction complexity of hierarchical bounding volumes can be too high, hence alternative approaches have been developed.

2.1.1.3 Grids

Perhaps the conceptually simplest of all acceleration data structures is the uniform grid, providing subdivision into fixed sized volumes throughout a given dimension of the domain resulting in predictable traversal computation. Construction is linear with respect to the number of elements being partitioned [ISP07], and hence are highly appropriate for fully dynamic data. Their shortcomings however are their non-adaptive nature, performing poorly under skewed data distributions due to their inability to subdivide regions with large quantities of data. Adaptive grids [KS97] provide a hierarchical approach to grid construction, nesting uniform grids inside non-empty nodes to provide a shallow adaptive hierarchies.

Fine grids can improve traversal performance, but for sparse data can incur high memory overheads, due to the number of empty cells still held in memory. Lagae and Dutré [LD08] present a hashed grid approach to minimise memory usage. Optimising grid construction for the GPU has also been investigated, for example in [KBS11] using a two-level hierarchical approach.

2.1.2 Specialised data structures

In addition to their general use in ray tracing, research has also focussed on specialising data structures to improve performance for particular ray tracing tasks. Djeu et al.[DKH09] utilise empty kD-tree nodes inside watertight meshes to represent occlusion volumes for fast shadow ray tests when coupled with a modified traversal procedure. Hunt and Mark [HM08b, HM08a] tailor axis aligned data structures to perform better for primary camera rays using a perspective transformation of the axis-aligned split planes.

Designing for hardware Orthogonal to much of the work discussed here, improving traversal performance by designing around bottlenecks of current hardware has seen notable improvements. Packet based ray tracing for coherent rays (such as rays for soft shadow tests or primary camera rays for visibility) has been studied for example in [BEL+07, GL10, GPSS07, WGBK07, RSH05] to leverage the similarities in trajectory of rays when computing intersection tests. However such approaches are not so suitable as ray coherence breaks down, requiring sorting and filtering to improve coherence [Tsa09] but is likely to provide little benefit to the incoherence found in global illumination and attempting to extract similar rays can restrict and complicate the rendering system architecture.

Making use of SIMD architecture of modern CPUs has been studied by Dammertz et al.[DHK08] and Wald et al.[WBB08], who increase the arity of the BVHs to match the number of SIMD lanes, moving away from binary trees towards shallower hierarchies that allow all child nodes to be intersected in parallel with a given ray.

2.2 Surface representation

Ray tracing provides a mechanism for computing visibility between objects and propagating light through free space. Once visibility has been determined, the appearance and shading of the object from the view point must be established. The interaction of light with surfaces is

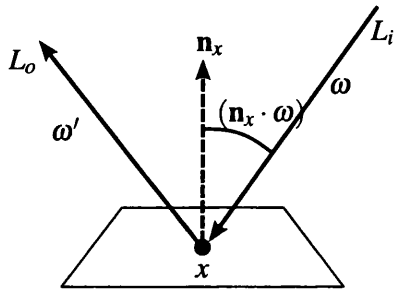


Figure 2.1: *Illustration of the geometric factors affecting the value of a BRDF* $f_r(x, \omega, \omega')$

complex, and dependent on a larger number of factors including properties of both the light arriving at the surface, and the composition and geometry of the surface itself at that precise location:

- Surface position,
- Relative viewing direction,
- Relative light source location,
- Wavelength of light,

For physically-based rendering, dependence on the lights wavelength is accounted for by evaluating the scattering of light for each wavelength independently, such that the scattered illumination is a multidimensional vector (of 3-dimensions in the case of RGB rendering). In practice the remaining factors can be divided into two distinct types:

- Positional variance, of the surface location,
- Directional variance, of the incident and exitant angles,

which was established by Blinn [BN76, Bli77], and form the foundation for many of today's rendering systems. The positional variance on a surface is responsible for adding detail, such as grains and knots in wood or patterns in wallpaper and is achieved through texture mapping.

Directional variance is arguably more significant for improving the realism of materials in rendering since in general the amount of light transmitted (and hence also absorbed) from a surface is dependent on the location of the viewer and light source relative to the local surface geometry. Functions describing this directionally variant scattering are known as Bi-directional Surface Distribution Functions, or BRDFs.

2.2.1 Bi-directional Surface Distribution Functions

Given a position x on a surface, a BRDF describes the proportion of light L_i arriving from an incident direction ω that is scattered towards an outgoing direction ω' as outgoing light L_o (Figure 2.1). BRDFs are commonly written as $f_r(x, \omega', \omega')$ (although other notation is also useful, and will be introduced later in the chapter). To provide accurate appearance models and adhere to physically-based rendering principles, there are two key properties that must hold for all plausible BRDFs:

- **Conservation of energy**: The total quantity of emitted light over the hemisphere at x, cannot exceed the quantity of incident light arriving at x.
- **Helmholtz reciprocity**: The value of the BRDF is constant with respect to the lights direction of travel, thus $f_r(x, \omega', \omega) = f_r(x, \omega, \omega')$.

Together these properties ensure that light is in equilibrium for all points in the scene, and as a result there exists a steady state for which a solution can be found.

2.2.1.1 Analytical models

A wide number of BRDFs can be described analytically, using functions to dictate the distribution of incident illumination over the hemisphere and can compute the BRDF for a given pair of input vectors. These aim to be computationally inexpensive, and are successful in approximating a wide range of interesting materials.

The simplest and most well known BRDF is the Lambertian reflection model, which scatters incident light equally in all directions over the hemisphere and is equal to $\frac{1}{\pi}$ for all ω and ω'. BRDFs with wide distributions over the hemisphere are referred to as low frequency BRDFs, and tend to produce soft indirect lighting.

In physically-based rendering, specular reflection and refraction from smooth conductors (eg: metals) and dielectrics (glass, diamond etc.) can be modelled using sets of *Fresnel equations*, describing the amount of light reflected (and hence refracted or conducted) for a given incident angle. This provides a more realistic appearance model than simple specular models that assume angle independent scattering.

Modelling rough surfaces Lambertian models are trivial to compute making them suited to high performance rendering. However their simplicity also neglects the more complex nature of real world materials. Perfectly smooth diffuse surfaces are uncommon in reality, so modelling surface roughness is highly useful in producing realistic BRDFs. Achieving this can be done through the use of microfacets, which model a surface made up of many small flat faces (the microfacets) with shading normals oriented around the geometric surface normal according to some statistical distribution. The facets are assumed to be significantly smaller than the area being illuminated, thus the appearance of the surface is considered to be the aggregate effect of many such facets. The roughness, and hence appearance of the material is dependent on a number of factors:

- **Geometry model**, describing the local lighting effects such as self shadowing and visual occlusion of the microfacets and interreflection.
- **Statistical distribution**, dictating the orientation of the microfacets which in turn provides the roughness of the surface,
- **Facet BRDF**, modelling the scattering of light for individual microfacets, as used when modelling a smooth surface.

Along with the distribution, allowing different scattering properties for individual microfacets makes it a highly flexible technique for modelling rough surfaces with diffuse or specular

scattering. Microfacets are commonly associated with specular interactions, however Oren and Nayar [ON94] introduced a diffuse model that assumes a Lambertian scattering model for each microfacet, described using a Gaussian distribution, which produces more realistic results at shallow incident angles.

Torrance and Sparrow [TS67] introduced one of the earliest microfacet models for specular reflection, which is independent of the distribution of the microfacets themselves, so long as a specular BRDF is used for the scattering behaviour. Thus it can be used as a basis for many materials and visual appearance models combined with different distributions and facet BRDFs. Blinn [Bli77] introduced a distribution that describes an exponential falloff of facet orientations from the geometric surface normal, better approximating illumination for non metallic surfaces and side lit geometry making it useful for rendering shiny plastics and varnishes.

Anisotropic BRDFs So far examples have been given of isotropic BRDFs, which are invariant with respect to the orientation of ω' and ω around the surface normal and tangent space. For example, the appearance of a sheet of plain paper when rotated about its normal does not change. Anisotropic models are dependent on this rotation around the surface normal, and thus are useful for modelling a range of materials such as brushed metals or fine woven fabrics. Most surfaces observed in the real world appear isotropic, like the sheet of paper, although they are unlikely to exhibit perfect isotropy due to small scale irregularities. For simplicity, they are commonly modelled using purely isotropic BRDF models in virtual environments.

Ward [War92] introduce an anisotropic BRDF based on an elliptical Gaussian function, whose distribution can be sampled from effectively (the significance of sampling outgoing directions from a surface using BRDF importance will be discussed in the following two chapters). Ashikhmin and Shirley [AS00] present an anisotropic microfacet distribution that is highly flexible, and can approximate both diffuse and specular materials of varying roughness's and anisotropies.

Measured data For more complex materials accurate analytical models are difficult to establish and limited functional light scattering models in turn limits the potential realism of rendering techniques. As a result data measured from real world materials (obtained using specialised equipment such as gonioreflectometers) can be employed to synthesise their appearance in virtual environments. An overview of such acquisition techniques can be found in [WLL+09]. Evaluating measured data usually requires interpolation methods [ZERB05] which can have a high impact on rendering performance due to the large quantity of data required to faithfully represent a BRDF. The realism that measured data can provide can however be unachievable via alternative methods.

Measured data for realistic BRDFs is available from, for example, the MERL BRDF database [Mit13] populated using the work by Matusik et al.[MPBM03] which requires a lookup table for regularly spaced BRDF measurements without the need for interpolation found in irregular measurement data.

2.2.2 Texture mapping

To provide surface detail, texture mapping can be applied that alters the effects of the BRDF scattering and is an essential part of realistic rendering. Such detail maps can be generated procedurally producing effects from simple functions like checkerboard or tiling patterns, or to synthesis complex natural materials such as marble, rock and snow using the pioneering work of Ken Perlin [Per85]. Like BRDFs, data can be acquired from the real world via photographs or computer generated images to provide improved realism.

Due to their fine detail, the use of texture maps can produce high variation in pixel values if appropriate filtering and anti-aliasing techniques are not used. Since both textures and image pixels have predetermined frequencies, based on the image or procedural function resolution of the texture, and the resolution of the rendered image respectively, this can cause severe artifacts. Luckily, both sample rates are known in advance, or can be approximated, and as a result the texture functions can be resampled to remove information of frequencies higher than the Nyquist limit of the image pixel filter [Ige99].

2.3 Light sources and emission

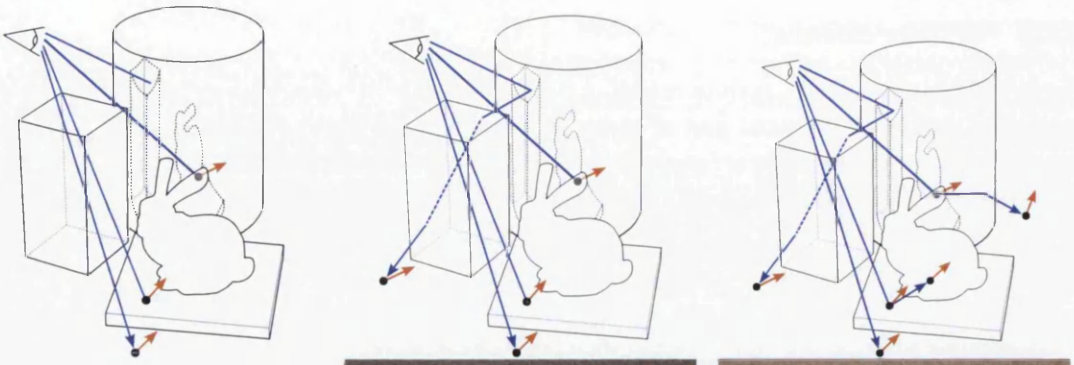
Illumination is a fundamental aspect of realistic imagery and without plausible representations, the lighting and the resulting realistic appearance of a scene can be weakened dramatically. Typically, the point light source is popular in computer graphics due to its simplicity, and due to its infinitesimal area and omni-directional distribution sampling and emission calculations are simple. Distant or directional light sources are also useful, in order to approximate light sources that are a long way from the scene geometry and have highly parallel rays.

However, neither approach is physically valid, since light sources in reality have finite area. As a result physically based rendering commonly uses area light sources, represented by geometric meshes with emissive properties. This makes illumination calculations more difficult, requiring integration over the surface of the light source in order to determine its contribution to the shading on a particular surface. Light sources that use complex meshes are however more computationally expensive, and as a result point and directional light sources are still useful in special cases to improve performance where their approximations are relatively accurate.

2.4 Light transport

Having identified some of the key components of physically based rendering using ray tracing, and motivation for their use, this section will discuss fundamental concepts and ideas to support an understanding of focus area of this thesis; light transport for global illumination.

Rigorous mathematical formulations of global illumination stem from its representation as a transport problem, where energy is propagated from one or more emitters, undergoing a number of interactions with the environment before being collected at a sensor. Mimicking the physics of light in this way is the most intuitive approach to tackle illumination of virtual scenes, accounting for light arriving directly at the camera sensor, and that which is scattered



(a) Ray casting
EDL

(b) Recursive ray tracing
ES^*DL

(b) Global illumination
$E(S|D)^*L$

Figure 2.2: *Illustration and examples of (a) Ray casting (b) Recursive ray tracing which accounts for reflection and refraction and (c) Global illumination which adds indirect illumination. Only the method of light transport has changed in each rendered image, illustrating its importance in realistic image synthesis.*

and provides a perception of the environment. Transport problems for global illumination exhibit strong similarities with those established in the fields of radiative heat transfer, acoustical engineering and particle physics problems, from which many concepts are borrowed (such as Monte Carlo integration) to aid computer graphics.

This section will introduce mathematical formulations for light transport and briefly discuss the complexities of solving light transport accurately. This leads into an appreciation for the need for Monte Carlo integration techniques, which is the focus of the next chapter, and the basis for the work presented in this thesis.

2.4.1 Ray tracing

Turner Whitted's [Whi80] recursive ray tracing builds upon ray casting from the camera to account for shading at surfaces using ray recursion (Figure 2.2). Unlike physical light transport, ray tracing is performed *backwards* from the camera to the light source(s) in the scene to reduce computation. In reality, only a small fraction of the light emitted actually hits a camera sensor

or the human retina, and hence is of interest in simulations, making backwards ray tracing a better approach. Casting a primary ray from the camera $r(x_0, \omega)$, provides a point x_1 on a surface visible from the camera (if one exists). From this, information about the surface geometry, BRDF and texture detail at that location can be provided for use during shading.

Secondary rays are then spawned from x_1 to include illumination effects, and can take one of three forms depending on the surface characteristics. Diffuse surfaces spawn *shadow rays* explicitly linking x_1 to one or more light sources in the scene, y to compute an estimate for the radiance arriving at x_1. For specular surfaces, explicit shadow rays will return zero radiance due to the narrow distribution of the BRDF, so instead secondary rays are cast along the angle of reflection or refraction recursively until reaching a light source or diffuse surface x_d, from which a radiance calculation can be performed. This sequence of points (known as *vertices*) forms a *path* through the scene. Typically, the number of vertices in each paths is controlled by terminating the recursion after a specified number of iterations to improve performance and prevent infinite recursion, for example in the case of light bouncing between a set of mirrors.

2.4.2 Path space

Light transport is discussed in terms of paths between the camera and the light sources, carrying importance or radiance depending on their origin. Due to the complexity of path construction and their characteristics it is useful to adopt a number of concise notations when discussing light transport in depth. *Path space* is the set of all possible paths Ω_* between any two points in the scene of any length:

$$\Omega_* = \bigcup_{k=1}^{\infty} \Omega_k$$

where k is the number of connections (rays) that form a path with $k+1$ vertices. In most cases, only paths formed between the camera and the light source are of concern, since paths outside this set do not contribute radiance to the image. Ray casting only deals with path space of length one (ie: Ω_1), connecting pixels on the camera sensor to visible points in the scene. Adding shadow rays extends this to paths in $k \leq 2$ ($\Omega_1 \cup \Omega_2$) where vertices are diffuse. Ray tracing additionally includes paths of arbitrary k when dealing with specular surfaces (due to its recursion) so does not fit neatly into this representation of path space. Instead, paths (and by extensions light transport algorithms) can be classified based on their interactions with the scene.

2.4.2.1 Path notation for surface interactions

It is often useful to know the type of interaction each vertex of the path has with the scene to provide a means of classifying paths based on the type of illumination they contribute to the image. Many algorithms are limited to specific types of light transport, or perform badly under certain conditions, it is useful to describe such paths succinctly. Heckbert [Hec90] introduced such a scheme based on regular expressions, using a single letter to represent each path vertex [1]. Using this notation the space of all paths that connect the camera with the light sources can

[1] For those unfamiliar with regular expressions, $A^?$ denotes zero or one occurrences of A, A^* denotes zero or more, and A^+ one or more occurrences. $(A|B)$ denotes the option of A or B

be expressed:

$$L(S|D)^*E \tag{2.1}$$

where L represents an intersection with a light source, Vertices that lie on the camera sensor are denoted E (for eye). Specular and diffuse interactions are classified by S and D respectively, describing whether the light carried along the path has undergone specular or diffuse scattering at this vertex.

The order of the L and E is also convenient for denoting a path's origin, and hence whether it propagates radiance from a light source, or importance from the camera. In practice $L(S|D)^*E$ and $E(S|D)^*L$ are equivalent with respect to path space. For recursive ray tracing, which extends from the camera the set of paths it handles can be denoted:

$$ES^*D^+L$$

Whitted's ray tracing model thus accumulates radiance from light sources visible to the camera or through specular reflection or refraction, ES^*L, and the path space ES^*DL known as *direct lighting* accounting for shading at the first diffuse vertex. In order to compute global illumination, this path space needs to be extended to encompass all possible paths in Equation 2.1. Figure 2.2 shows the effect of rendering these additional paths, alongside an illustration and their regular expression notation. Diffuse inter-reflection is responsible for colour bleeding, and caustics can add high energy features, both of which are crucial for purveying realism.

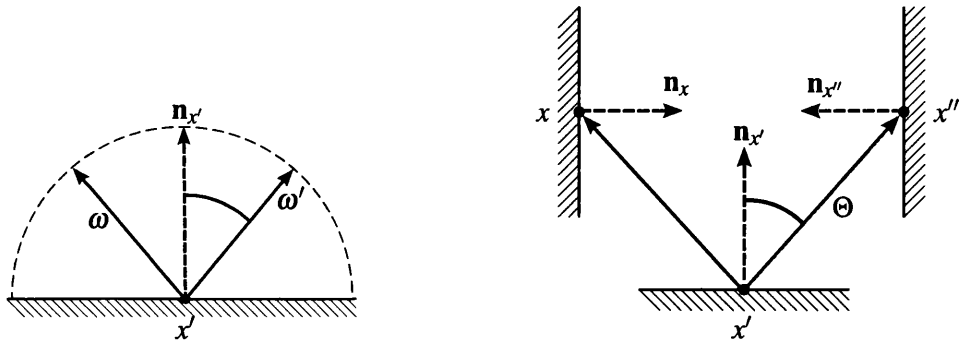
2.4.3 Global illumination as an integration problem

The original ray tracing technique as proposed by Whitted deals strictly with perfectly Lambertian reflectors in conjunction with purely specular surfaces and point light sources. This presents a significant simplification over the need to solve integrals, instead evaluating illumination from a number of discrete points in the scene. Extending this discretisation approach to handle soft phenomenon (such as shadows from area light sources) results in banding and image artefacts if carried out naïvely by sampling discrete points in the integral. To enable the use of more realistic environments, more advanced techniques are required that can handle this additional complexity efficiently without introducing artefacts.

Cook et al.[CPC84], applied random sampling to ray trace a variety of effects including soft shadows, motion blur and depth of field, allowing such phenomenon to be solved in their full integral form. Applied to ray tracing this still only accounts for directly visible effects, but treats ray tracing as a set of interrelated problems that together form a higher dimensional integral, as opposed to a series of individual integrals of low dimension which cannot effectively be solved on their own.

Kajiya [Kaj86] presented a recursive integral equation from radiative heat transfer literature in a form suitable for rendering global illumination. This extends the idea of Cook's stochastic ray tracing, allowing the inclusion of indirect lighting; light scattered towards visible surfaces from other surfaces in the scene. The modern form of Kajiya's *rendering equation* is the theoretical basis of today's global illumination research and is commonly expressed as:

$$L(x',\omega) = L_e(x',\omega) + \int_{\Omega_{x'}} L(x',\omega') f_r(x',\omega,\omega') \cos(\mathbf{n}_{x'},\omega') \mathrm{d}\omega' \tag{2.2}$$



(a) Hemispherical formulation

(b) Surface area (or three-point) formulation

Figure 2.3: *Terms used in the two formulations of the rendering equation and light transport theory. Terms for the hemispherical formulation (Equation 2.2) using unit vectors ω and ω', and the surface area (or three-point) formulation used in Equation 2.3.*

which describes the total radiance L scattered in direction ω from a point x' on a surface in the scene (Figure 2.3a) as an integral of incoming radiance L from ω' over the unit hemisphere at x', $\Omega_{x'}$. This also accounts for the emitted radiance L_e originating from x' itself. Here the integral is . This equation assumes no participating media is present in the scene and hence radiance is not lost when travelling between surfaces, which is commonly the case. The rendering equation can be generalised to account for the scattering of light between surfaces and a concise discussion of participating media can be found in [Jen01] in the context of photon mapping.

Note that this formulation of the rendering equation integrates over the hemisphere at x', which is convenient when dealing with ray tracing operations based on sampling using BRDF distributions. Alternatively, radiance can be integrated with respect to the surface area of objects visible from x', which is convenient in some circumstances when discussing light transport between surface patches or from arbitrary points in the scene (such as light sources). Given two points x and x'' (Figure 2.3b), the rendering equation at x' can be rewritten:

$$L(x' \to x) = L_e(x' \to x) + \int_S L(x'' \to x') f_r(x'' \to x' \to x) G(x'' \leftrightarrow x') V(x'' \leftrightarrow x') \mathrm{d}A_{x''} \quad (2.3)$$

where S is the set of all surfaces in the scene. $L(x' \to x)$ is the exitant radiance from x' to x. This is commonly known as the *three-point* or *area formulation* of the rendering equation.

The geometry term $G(x'' \leftrightarrow x')$ appears due to the need to transform the integral from solid angle to unit area measure, accounting for the grazing angles of exitance at x'' and angle of incidence at x':

$$G(x'' \leftrightarrow x') = \frac{\cos(\mathbf{n}_{x'}, \Theta)\cos(\mathbf{n}_{x''}, -\Theta)}{\|x' \to x''\|^2} \quad (2.4)$$

where $\|x' \to x''\|$ is the magnitude of the vector from x' to x'', and Θ is its normalised unit vector:

$$\Theta = \frac{x' \to x''}{\|x' \to x''\|}$$

The binary visibility function $V(x'' \leftrightarrow x')$ represents the mutual visibility between two points and ensures the propagation of radiance between x'' and x':

$$V(x'' \rightarrow x') = \begin{cases} 1 & \text{if } r(x', \Theta) = x'', \\ 0 & \text{otherwise.} \end{cases}$$

In the solid angle formulation, this is implicit since vectors towards ω and ω' will only intersect visible geometry by definition of the ray tracing functions $r(x', \omega)$ and $r(x', \omega')$ respectively. Equations 2.2 and 2.3 are equivalent, and provide useful formulations of the light transport problem, both of which will be used throughout this thesis.

The recursive nature of light transport becomes evident when radiance from surfaces visible over the hemisphere at x' must be included in $L(x' \rightarrow x)$. Incidentally, radiance scattered from x itself towards x' (or via x'' to x') can also contribute to $L(x' \rightarrow x)$. Hence in theory the integral over all surfaces in S must be evaluated, resulting in an infinitely recursive integral. Correctly estimating this indirect illumination accounts for the high complexity of global illumination. The rendering equation can therefore provide a theoretical framework for all physically-based light transport algorithms. The rendering equation only describes how radiance travels around the scene. To obtain an accurate appearance of a scene, the *importance* of a region or point with respect to the virtual camera needs to be considered in addition to the propagated radiance. Importance W, is the dual quantity of radiance, hence substituting W for L yields an equation describing the flow of importance through a point:

$$\begin{aligned} W(x, \omega) &= W_e(x', \omega) + \int_{\Omega_{x'}} W(x', \omega') f_r(x', \omega, \omega') \cos(\mathbf{n}_{x'}, \omega') \mathrm{d}\omega' \\ &= W_e(x' \rightarrow x) + \int_S W(x'' \rightarrow x') f_r(x'' \rightarrow x' \rightarrow x) G(x'' \leftrightarrow x') V(x'' \leftrightarrow x') \mathrm{d}A_{x''} \end{aligned} \tag{2.5}$$

By evaluating the rendering equation and accounting for camera importance, the measured radiance M for any path can be computed. The *measurement equation* for a pixel p is of more interest in global illumination, obtained by:

$$M_p = \int W(x', \omega) L(x', \omega') f_r(x', \omega, \omega') \cos(\mathbf{n}_{x'}, \omega') \mathrm{d}A_{x'} \mathrm{d}\omega' \tag{2.6}$$

or using the area formulation:

$$R_p = \int W(x \rightarrow x') L(x'' \rightarrow x') G(x' \leftrightarrow x'') \mathrm{d}A_x \mathrm{d}A_{x'} \tag{2.7}$$

with the additional constraint that x is a point on the camera sensor inside the region of pixel p, with x' and x'' in S.

In relation to our regular expression path notation for surface interactions, it is generally the case that paths of the form $E...L$ propagate importance using Equation 2.5 and paths originating at a light source, $L...E$, transport radiance according to the rendering equation. In practice the mechanisms behind both are the same, but this distinction becomes important when discussing the operation of light transport algorithms and their characteristics for evaluating complex light transport.

Conclusion This chapter has provided a brief overview of the fundamental components of a ray based renderer, and the accurate representation of virtual environments. A mathematical model of scene objects, and the means to solve visibility and occlusion queries between them using rays has been outlined. This included detailed discussion into the need for acceleration data structures for ray casting, providing much needed performance improvements over naïve approaches.

The representation of materials using BRDFs and importance of their efficiency was presented along with details of the implementations used to produce images in this thesis. The application of texture maps has been mentioned as a convenient method of adding the fine detail necessary for realistic renders. Further discussion of common light sources and camera models followed, as a means of modelling and measuring light emission in the scene.

This chapter concluded with a discussion of light transport using ray tracing, and the need for recursive evaluation to tackle the complex problem of indirect lighting. Fundamental equations in the theory of light transport such as the rendering equation were introduced, and notation used throughout the remainder of the thesis was established including the concept of path space and the concise regular expression notation used to describe it.

With this theory of light transport in mind, the next Chapter will discuss the mathematical background to Monte Carlo integration and stochastic sampling, fundamental to state of the art techniques for global illumination.

Chapter 3

Monte Carlo Integration and Rendering

Contents

3.1 Probability theory **29**
3.1.1 Probability distributions and density functions 30
3.1.2 Expected values and variance 31
3.2 Basic Monte Carlo integration **31**
3.3 Variance reduction techniques **34**
3.3.1 Importance sampling 34
3.3.2 Generating continuous random variables 36
3.3.3 Russian roulette and splitting 37
3.3.4 Adaptive sample generation 38
3.3.5 Markov-Chain Monte Carlo 39
3.3.6 Additional techniques 39

This chapter provides an introduction to Monte Carlo integration, and discusses its use and importance in solving the rendering equation for global illumination. It will provide a foundation for understanding the complexity of high dimensional problems, and motivation for applying probabilistic and stochastic (random) theory in order to provide efficient solutions. Several approaches to improve sampling distributions and reduce variance for stochastic methods will be discussed, along with a discussion on more advanced adaptive sampling techniques.

3.1 Probability theory

Probability theory, as a basis for statistics, is an essential tool for understanding and modelling the behaviour of large and complex systems. From the discussion of physical models of illumination in Chapter 1, light transport is one such a system and probabilistic methods are crucial in

reducing the complexity involved in simulating it. Building upon probabilistic methods allows a good estimate to the solution of a complete system to be obtained with only partial knowledge of its state. Computationally, this means that accurate models of complex systems can be acquired more cheaply than brute force approaches, providing such stochastic and probabilistic techniques are used appropriately.

Random variables are used in mathematics to describe the possible outcomes of a stochastic experiment, based on the occurrence of random events. A random variable X may take on any one of a number of possible values, each with an associated probability $p(X)$ of being chosen, where $0 < p(X) < 1$. Such variables can be either discrete (whereby the number of possible outcomes is countable), or continuous where it could take on any one of an infinite number of values within some specified interval or intervals, denoted Ω [1]. In turn such interval(s) can be finite or infinite.

In computer graphics we deal with both discrete and continuous random variables usually within finite intervals. Take the example of selecting a point on the surface of an object in a scene. There are countably many objects hence a finite number of outcomes, but for each object there are infinitely many points on its surface that can be chosen [2].

3.1.1 Probability distributions and density functions

Values for random variables are drawn based on a *probability distribution function*, that describes the likelihood of the random variable to take on a given value. The term *probability density function* generally applies to continuous domains and *probability mass function* is used in the discrete case. In this thesis the term *probability density function* (shortened to *PDF*) will be preferred since the continuous case is more often discussed, although they can be used synonymously for purposes of discussion. A valid probability function ensures that $\forall x \in \Omega, p(X) > 0$ and $p(\Omega) = 1$. That is, there must be a non-zero probability of choosing any value within Ω, and the probability of sampling all points in Ω must equal 1 (guaranteeing that any sample drawn from p lies in Ω). For continuous random variables a probability density function is used to describe the probability Pr, that the random variable X will lie within a given region of that domain:

$$Pr[a \leq X \leq b] = \int_a^b p(X)\mathrm{d}X \tag{3.1}$$

for some $a, b \in \Omega$. Since in $Pr[a \leq X \leq b] = \int_a^a p(X)\mathrm{d}X$ mple has a zero probability of being drawn individually, and thus

Using random variables drawn from such distributions has some useful properties. Given a function $f(X)$ dependent on X, properties that hold for X subsequently hold for $f(X)$, since its value is dictated by the underlying distribution of X given by $p(X)$. Often the result of a function $f(X)$ over some domain is of more interest than the value of X itself, since they provide the solution to the integrand; its *expected value.*

[1] Recall that in the previous chapter, $\Omega_{x'}$ was defined as the area over the hemisphere of a point x on a surface, since $\Omega_{x'}$ is an interval within the integrand of interest over which we can draw samples

[2] In practice, since limited precision data types are used there are in fact a finite number of values that can be represented, but as this number is large and implementation dependent it is convenient to treat it as the continuous case.

3.1.2 Expected values and variance

The *expected value* or mean of a random variable is the weighted average of all possible values of X. In the discrete case, the expected value of a random variable $f(X)$ with values of X in an arbitrary domain Ω can be obtained by the summation of all possible $f(X)$ weighted by their probability, $p(X)$. For continuous random variables, the expected value becomes an integral over Ω:

$$E[f(X)] = \int_{\Omega} f(X)p(X)\mathrm{d}X. \tag{3.2}$$

A function used to compute $E[f(X)]$ over Ω is said to be an estimator of $f(X)$. *Variance* is a useful means of quantifying the error of an estimator with respect to the true solution. The variance can be described as the expected deviation of a function or variable, from its expected value:

$$V[f(X)] = E\left[(f(X) - E[f(X)])^2\right]$$

or alternatively, variance can be expressed as the expected value of the square, minus the square of the expected value:

$$V[f(X)] = E[f(X)^2] - E[f(X)]^2. \tag{3.3}$$

Amongst other properties this holds since, for a set of random variables the expected value of their sum is equal to the sum of their expected values:

$$E\left[\sum_{i=1}^{N} f(X_i)\right] = \sum_{i=1}^{N} E[f(X_i)] \tag{3.4}$$

and, provided the random variables are drawn independently, the same identity holds for variance:

$$V\left[\sum_{i=1}^{N} f(X_i)\right] = \sum_{i=1}^{N} V[f(X_i)]. \tag{3.5}$$

These properties are useful when combining expected values from independent estimators, since any estimator with N samples, can be viewed as a combination of N individual estimators each with a single sample. In practice, this is one of the key concepts for the success of stochastic integration techniques in rendering, providing a running expected value useful for gradual refinement of the solution, and allowing flexible implementations that can be tailored to massively parallel hardware running multiple independent estimators that can later be combined.

3.2 Basic Monte Carlo integration

For integrals where analytical solutions are not available or difficult to establish, or where the integrand can only be easily computed at certain points, numerical integration techniques can be used to compute an approximate solution to such an integral over a domain Ω:

$$I = \int_{\Omega} f(X)\mathrm{d}X,$$

for which there is a function f that can be evaluated for a random variable $X \in \Omega$. Where Ω is of one-dimension, analytical methods may be available or techniques derived from the interpolation of functions across subintervals may be applicable. As the dimensionality of the domain increases, such methods can be applied by breaking down the integral in to several one-dimensional sub-integrals. However such methods quickly lose their efficiency at higher dimensions since the number of evaluations of f required increases exponentially.

Monte Carlo techniques are a practical alternative for solving problems in higher-dimensions, evaluating f at a number of random points within Ω, to provide an expected value of I. Supposing we have a PDF $p(X)$ over Ω, where $\forall X \in \Omega$, $p(X) > 0$. An approximation to I can be obtained using a basic Monte Carlo integrator by evaluating N independent samples drawn from the PDF:

$$\langle I \rangle_N = E[I] = \frac{1}{N} \sum_{i=1}^{N} \frac{f(X_i)}{p(X_i)} \tag{3.6}$$

where $X_1 ... X_N$ are independent random variables distributed in Ω. If samples are drawn uniformly over Ω then $p(X)$ is constant, $\forall X \in \Omega$, reducing it to the mean of $f(X)$. Due to the law of large numbers, $\langle I \rangle$ is an unbiased approximation to I converging towards I as N increases:

$$\lim_{N \to \infty} \langle I \rangle_N = I. \tag{3.7}$$

As mentioned earlier with respect to $f(X)$, the estimator $\langle I \rangle$ itself is also dependent on the underlying distribution of X since it is also a function of X. Identities for expected variance established for random variables X_i (Equations 3.4 and 3.5) therefore also hold for Monte Carlo estimators:

$$V[\langle I \rangle_N] = V\left[\frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{p(x_i)}\right] = \frac{1}{N} V\left[\frac{f(x_i)}{p(x_i)}\right].$$

The standard deviation σ is also known as the *root-mean squared error* (RMS error), proportional to the error of the estimator, and is generally more meaningful than variance, since it is expressed in the same measurement units as $f(X)$. Calculating the RMS error of the estimator is a useful quantitative tool for comparing the effectiveness of different algorithms to evaluate an integral if the exact solution is known. The standard deviation of $\langle I \rangle$ can be obtained from its variance:

$$\sigma[\langle I \rangle_N] = \frac{1}{\sqrt{N}} \sigma\left[\frac{f(x_i)}{p(x_i)}\right]. \tag{3.8}$$

As N increases, variance decreases linearly with N and the error of the estimator decreases with $\sqrt{N}$, providing Monte Carlo methods with a converge rate of $O(\sqrt{N})$. Given the true solution of the integral I, the error of a Monte Carlo estimator over I is given by:

$$Error(\langle I \rangle_N) = I - \langle I \rangle_N.$$

Taking the expected value of the error provides the *bias* β associated with the estimator. This is the difference between the expected value of the estimator and the true solution of the integrand:

$$\beta[\langle I \rangle_N] = E[\langle I \rangle_N - I],$$

thus even as $\lim_{N\to\infty}$, the bias of the estimator will be non-zero, $\beta[\langle I\rangle_N] > 0$. Monte Carlo methods are considered *unbiased* since they do not introduce systemic error, only statistical error due to insufficient N, thus under sampling the domain. For a Monte Carlo estimator, or indeed any unbiased estimator it holds that:

$$\forall N : E[\langle I\rangle_N] = I.$$

Estimating the error of an unbiased estimator is significantly simpler than in the case of a biased estimator since any discrepancy between $\langle I\rangle_N - I$ is the result of error alone, and is trivially computed using $V[\langle I\rangle_N]$. In the case of biased estimators, this discrepancy is an unknown combination of error and bias, $V[\langle I\rangle_N] + \beta[\langle I\rangle_N]$ which is often difficult to estimate.

A crucial concept for global illumination is the use of *consistent* estimators. An estimator is said to be *consistent* if the variance and bias of the estimator goes to zero in the sample limit, with respect to the true solution of the integral, ie:

$$\lim_{N\to\infty} V[\langle I\rangle_N] = \lim_{N\to\infty} \beta[\langle I\rangle_N] = 0,$$

thus ensuring that for a given integral, the estimator will always produce the true result:

$$\lim_{N\to\infty} \langle I\rangle_N = I.$$

Thus unbiased estimators that are consistent, are *strongly consistent.* Biased estimators can be *weakly consistent* if their variance goes to zero, and they converge to some quantity $\theta \approx I$:

$$\lim_{N\to\infty} \langle I\rangle_N = I + \lim_{N\to\infty} \beta[\langle I\rangle_N] = \theta. \tag{3.9}$$

The consequences of these estimator properties and the impact they have on designing algorithms for rendering will be discussed in Chapter 4.

The efficiency of an estimator is another critical concept to bear in mind, since σ and N alone do not dictate the practical effectiveness of the estimator unless the cost of sample evaluation is also considered. The efficiency of an estimator can therefore be defined as:

$$\varepsilon[\langle I\rangle] = \frac{1}{\sigma^2\tau}$$

where τ is the average cost of evaluating a single sample. For high dimensional integrals it is difficult to reliably determine τ for a given estimator, except by empirical means. By running estimators for a given length of time and assessing their error (and bias if applicable), the efficiency of estimators can be reliably compared. The efficiency and RMSE measures will be used to evaluate some of the techniques proposed in this thesis empirically with comparisons to alternative estimators.

Advantages Monte Carlo methods have several advantages for complex problems. They can be applied widely since they have few restrictions. In their most basic form they require only two operations. First, a valid probability density function $p(X)$ from which to draw random

samples from the domain. Second, the ability to evaluate the function of interest $f(X)$ for such samples.

The convergence of Monte Carlo methods, although slow compared to other numerical integration techniques, is independent of the dimensionality or smoothness of the integral. Singularities and discontinuities in the integral can be a problematic for many methods even in lower dimensions, but are handled elegantly by Monte Carlo integration. Integrals of arbitrary dimensions can be handled elegantly by simply extending sampling into those dimensions using an appropriate PDF, including recursive and infinite dimensional integrals of which light transport is an example. Furthermore, the estimator can provide expected values for such problems with no lower bound on the number of samples that need to be evaluated, unlike alternative techniques in which the minimum number of samples grows exponentially with dimension. This is thanks to the property described in Equation 3.4, allowing many estimators to be combined each with an arbitrary numbers of samples. This has further practical implications, allowing multiple estimators to be run independently, providing excellent scalability on parallel architectures.

3.3 Variance reduction techniques

The elegance and generality of Monte Carlo methods is tainted by its poor convergence when compared to numerical methods used for lower dimensional problems. To reduce the error of the estimator by half, the number of samples forming the estimate must quadruple. Evaluating the function $f(X)$ over the random variables often consumes the majority of the computation time, hence finding alternative techniques to improve the error of the estimator without additional samples has been well studied. Crucially, there are a number of general techniques that can reduce the variance of Monte Carlo methods without the computational cost of evaluating additional samples. In this section we will provide an overview of some techniques that have become popular in rendering and how they are applied to light transport since they will be referred to throughout the thesis.

3.3.1 Importance sampling

It is evident that the error of the Monte Carlo integrator relies heavily on the probability function $p(X)$. Importance sampling is a technique that attempts to choose a probability function intelligently so as to reduce error in the estimator. Any convenient PDF can be chosen provided $p(X) > 0$ when $f(X) \neq 0$, and with the obvious condition that it must be feasible to sample from $p(X)$ for any valid $X \in \Omega$. In many cases, especially in graphics, some information is known about the integrand and the shape of $f(X)$ prior to sampling; this is known as *informed* Monte Carlo. Conversely, under *blind* Monte Carlo methods no information is available to aid the sampling process and so a uniform density for $p(X)$ is the only feasible solution, since a bad choice of $p(X)$ can in fact increase variance.

Often, it is unfavourable to draw random variables from a uniform distribution over Ω, since the value of $f(X)$ could be zero for large regions of the integrand, especially in high dimensional cases. Conversely, exceptionally high values (relative to the true sample mean) may

be present in only small regions of the domain, thus sampling Ω poorly can lead to increased variance.

An optimal choice of p would ensure it follows the value of $f(X)$ throughout the domain, such that $\frac{f(X)}{p(X)}$ is constant for all X, and equal to I. Since this is impossible without knowing I itself, a function $g(X)$ may exists that is an approximation to $f(X)$ can be chosen based on available information, accounting for known factors of $f(X)$. Assuming $g(x) \propto f(X)$ is accurate, setting $p(X) \propto g(X)$ will reduce variance in the estimator compared to a uniform distribution.

Importance sampling is one of the fundamental variance reduction techniques in Monte Carlo rendering, since a wealth of partial information is available to improve the sample distribution especially in the case of highly glossy BRDFs and Dirac distributed functions. Further discussion and insight into importance sampling and related techniques can be found in [OZ98].

Defensive importance sampling Despite the effectiveness of importance sampling to concentrate samples towards areas of potentially large $f(X)$, it can be difficult to find an appropriate $g(X)$ such that $g(X) \propto f(X)$ throughout the entire domain. If $g(X)$ has regions of low density $\frac{f(X)}{p(X)}$ is unbounded since $p(X)$ can become arbitrarily close to zero, and hence the variance of the estimator is also unbounded producing noise spikes. To eliminate such problematic cases, Hesterberg [Hes97] suggested instead to use a mixture density combining importance sampling with a more uniform sampling distribution. Mixing with a uniform distribution ensures that the value of $p(X)$ is clamped to some minimum value above zero for all X, bounding the variance. The resulting mixture distribution significantly reduces the general effectiveness of importance sampling, since in regions where $g(X) \propto f(X)$, the mixture of $g(X)$ with the uniform distribution will not be. Thus in the general case defensive importance sampling actually increases the overall variance, despite bounding the variance of individual samples.

Resampled importance sampling Instead of mixing with a uniform distribution, an attempt can be made to compute a better representation of the integral, and sample according to this approximation. Supposing M samples are taken using some importance sampled PDF $p(X)$. A weight for each M can be obtained by partially evaluating $f(X)$ for each sample and a new discrete PDF, $q(X)$, constructed (with M possible outcomes) that is an approximation of $f(X)$. Drawing and evaluating N full samples from $q(X)$ can reduce variance if $q(X)$ is a better approximation to $f(X)$ than $p(X)$, which is the case as M tends to infinity. Obviously, this is only feasible if the partial $q(X)$ can be obtained with good efficiency. In illumination, this technique has been applied to direct lighting [TCE05] whereby the costly visibility term is ignored to partially evaluate the M initial samples. This approach can reduce variance, but is limited with respect to implicit lighting since the costly visibility term is required to produce M, and it becomes degraded for explicit lighting when occlusion is a significant cause of variance as it is ignored when constructing $q(X)$. Whilst providing good variance reduction for large M, sample generation can be prohibitively costly and just as standard importance sampling produces spikes in the integrand, this approach can suffer the same shortcomings.

3.3.1.1 Multiple importance sampling

Importance sampling can be effective in generating samples from a single function. Often, complex integrals are the product of multiple functions ($g(X)$ and $h(X)$), a single distribution function representing $g(X) \circ h(X)$, the composition of such functions may be difficult to obtain. In computer graphics, to determine the direct illumination at a surface during rendering, samples can be distributed according to the surface BRDF and cosine functions at the surface, or according to the flux distribution of the light source. Without being able to produce a single representative distribution, sampling must be done with respect to one of these functions, $g(X)$ or $h(X)$. Arbitrarily choosing one of these distributions may lead to a bad sample distribution, which as discussed above can increase the error of the estimator compared to using the uniform distribution.

3.3.2 Generating continuous random variables

Whilst importance sampling ensures that samples are distributed with respect to the value of the integral, the underlying distribution of the random variables X themselves must also be considered.

In practice, probability density functions can be viewed as a functional black box. Commonly, a random number $r \in \mathbb{R}$ in the interval $[0, 1)$ is provided to the PDF, which is mapped to a random variable $X \in \Omega$ from which $f(X)$ can be evaluated. Generating numbers in $[0, 1)$ is convenient since it simplifies the mapping from r to X and can be applied to arbitrary domains.

Good quality generation of r is therefore important to ensure well distributed samples. Uniform generation is undesirable since even though variance is minimal, it can result in artefacts that cannot easily be detected numerically. If points are generated too close together the variance reduction power of samples to the estimator is reduced, since coverage of the domain is inefficient. Meanwhile, no large regions of the domain should remain unsampled, else the resulting estimator will be subject to systematic bias. Mathematically, the quality of a distribution is known as it's *discrepancy* with low-discrepancy sampling being more desirable than high.

Stratified sampling When a known number samples are desired prior to sampling, *stratification* can be employed to compute well distributed sample batches. Stratified sampling divides the domain into distinct cells with equal area (or volume for higher dimensions), known as *strata*. One sample is allocated to each strata in order to distribute the samples evenly across the domain to reduce the chance of missing important features of the integrand. Each sample is *jittered* (offset by some random amount) so as to reduce the artefacts associated with uniform sampling.

In graphics, stratified sampling is commonly applied to 2D problems such as pixel anti-aliasing or lens sampling for depth of field although it can be applied to higher dimensions by decomposition. Despite the use of strata, samples can still exhibit a degree of clumping at cell boundaries. This is improved by the use of Latin hypercube sampling and shuffling, but loses its practicality and effectiveness for larger sample counts.

Generating good samples is a complex and well studied topic, thus only an introduction to relevant aspects will be provided here. Further discussion on sampling strategies can be found in general rendering texts such as [PH10].

Quasi-Monte Carlo Quasi-Monte Carlo methods use quasi-random distributions which resemble random distributions, but additionally provide sample stratification without the need for explicit strata. Quasi-random sequences are fully deterministic, but provide desirable properties without introducing the uniform placement that results in banding. The benefit of such sequences is that the number of samples generated does not affect the stratification of the samples. For example with stratified sampling one batch of sixteen samples will provide a far better distribution than sixteen batches of individual samples, which is analogous to sampling using pseudo-random numbers.

The van der Corput or Halton sequences are often used in graphics, since the number of samples is usually not known or is too large to efficiently apply an explicit strata based approach. Despite their low discrepancy, care must be taken when using sequences in parallel or in neighbouring regions due to their deterministic and correlated nature. The use of random shuffling to scramble sample sequences such as that introduced by Kollig and Keller [KK02] provides a practical method of alleviating such problems. This approach is highly useful for multi-dimensional sampling and is used to generate sample patterns for the rendering examples in this thesis. Contrary to such benefits, obtaining error estimates becomes difficult since it is no longer a purely stochastic process. By carefully producing scrambled sequences, both the desirable quasi-random distribution properties and benefits of stochastic processes such as the ability to combine independent estimators can be maintained.

3.3.3 Russian roulette and splitting

Russian roulette and splitting are related techniques that increase the efficiency of Monte Carlo estimators by increasing the contribution of each evaluated sample to the final estimate, while maintaining an unbiased estimator. They were introduced to computer graphics by Arvo and Kirk [AK90]. Russian roulette trades computational cost for increased variance by skipping the evaluation of samples that are likely to have low contribution to the estimator. A threshold q is chosen and dictates the probability that an estimate for the solution will not be evaluated for a particular sample. Drawing a canonical random variable $\xi \in [0,1]^{\mathbb{R}}$ allows us to decide the value of the single sample estimator:

$$\langle I \rangle = \begin{cases} \langle I \rangle / q & \text{if } \gamma < q \\ 0 & \text{otherwise} \end{cases}$$

In the context of Monte Carlo rendering, the ray casting calculations for extending a random walk is the most costly operation [3]. The current importance of a camera path can provide a tentative estimate for that samples contribution. Terminating a path stochastically using this

[3] A random walk is a mathematical formalisation for a path constricted by taking random steps through a domain. In this case, the process of casting rays stochastically from one surface point to another can be thought of as a random walk through the scene.

potential estimate as γ can save additional costly visibility queries. From a theoretical standpoint, Russian roulette also allows a practical approach for random walks in infinite integrals such as the rendering equation. Applying Russian roulette at each vertex of the random walk and weighting any subsequent vertices accordingly provides an unbiased method of terminating random walks through infinite dimensions. Efficiency is also addressed, as longer paths become less likely so the algorithm concentrates on shorter walks where $f(X)$ is larger on average.

All samples evaluated for an estimator have similar computational cost, and hence the efficiency of the estimator is improved if time is spent on samples where $f(X)$ is large which have a high capacity to reduce variance as we have seen with importance sampling. Using Russian roulette provides an estimator with higher per sample variance, but on average the time to evaluate the integral is reduced given a reasonable value for γ.

Splitting, on the other hand creates additional samples in regions of the integrand deemed important, thus each step of the random walk is 'split' to evaluate multiple estimators for a single term of the random walk, increasing sample coverage over a particular region of the domain:

$$\langle I_v \rangle = \frac{1}{N} \sum_{i=1}^{N} \langle I_v \rangle_i$$

where $\langle I_v \rangle_i$ are independent estimators for a term v in the random walk $\langle I \rangle_1 + ... + \langle I \rangle_N$. The caveat with this approach is that there is no guarantee that efficiency will be improved, since splitting and using a number of estimators M where $M \ll N$ may also reduce variance to the same degree as splitting using N estimators, resulting in oversampling. In rendering, splitting is more commonly applied to special cases, such as specular reflection and refraction where the number of splitting possibilities is small, but the variance reduction may be significant. However, care must be taken since successive splitting can lead to a considerable increase in the number of samples required for even a small part of the integral to be evaluated, potentially with minimal contribution.

3.3.4 Adaptive sample generation

Importance sampling relies on finding a PDF a-priori that resembles the ideal PDF proportional to $f(X); \forall x \in \Omega$, in an attempt to place samples intelligently to reduce variance. Adaptive sampling aims to reduce variance directly by concentrating samples in regions that are considered important using previously evaluated samples. Commonly, variance is used as a metric to ensure sampling is increased for regions that exhibit high variance and hence where the error of estimator is greatest. In contrast to importance sampling, no information about the integrand is needed before sampling begins, since the PDF used to distribute samples can be generated and adapted on the fly during sampling. In the case of blind Monte Carlo or when no known PDF is an appropriate approximation to f, adaptive sampling can greatly reduce variance. Producing an adaptive sampling scheme that does not introduce bias is however challenging, since the entire domain still needs to be sampled to ensure the convergence of $\langle I \rangle$ to I. To apply adaptive sampling, often the domain needs to be divided into discrete regions, from which a probability distribution function can be obtained. Sampling from a discrete PDF over a continuous function can lead to increased variance if the discretisation is coarse or the integral contains sharp

discontinuities that are not captured by the discrete PDF. Furthermore, the measure of variance obtained during sampling does not necessarily reflect the true error present in the estimator. Absolute variance with respect to the solution I is still unknown, since variance and error metrics can only be calculated relative to the samples observed so far that make up $\langle I \rangle$. If part of the integrand has not yet been sampled, variance in that particular region may be low, hence it is likely to be avoided by adaptive sampling techniques. For this reason, adaptive sampling is often mixed with uniform or importance sampling to ensure improved domain coverage and convergence.

3.3.5 Markov-Chain Monte Carlo

Markov-Chain Monte Carlo (MCMC) methods are a collection of techniques that generate samples based on the construction of Markov chains, in which the generation of a new sample is dependent solely on the previous sample. Each new sample is weighted such that samples are distributed according to the target distribution. In the case of Monte Carlo methods this describes the target distribution in the limit, with a density proportional to $f(X)$ everywhere, hence more samples are distributed where $f(X)$ is large. Particularly, in MCMC methods such a distribution can be reached without any knowledge of $f(X)$ or its PDF, unlike importance sampling and control variates which rely on finding approximations to $f(X)$ before sampling.

Initially, a sample is generated to evaluate a random walk through the domain. Each subsequent sample is then generated by the mutation of the random variables that make up the previous sample; a process common to Markov chain methods. In complex integrals, this permits local exploration of the integral through important regions without having to generate samples stochastically, but by gradual change. By repeatedly applying mutation based operations over the previous sample alone, a chain of samples can be evaluated that tend towards parts of the integral that contribute significantly to the estimator, building up a distribution of samples that are generated according to f without knowing the value of f anywhere, meanwhile avoiding the need for approximations to the integral.

Markov-chain methods perform well for difficult integrals, especially where f is zero for large areas of the domain. Once an important region is found, the use of sample mutation can be effective for local exploration, where traditional Monte Carlo methods may struggle due to the low probability of sample placement in those regions. Additionally, MCMC methods are orthogonal to many of the other variance reduction methods discussed here such as (multiple) importance sampling and Russian roulette, with the exception of sample stratification which is dictated by the mutation strategy.

Several MCMC methods have been applied to rendering global illumination, and have proven effective under difficult lighting scenarios. An overview of the resulting algorithms will be discussed in the next chapter with a focus on their performance for global illumination.

3.3.6 Additional techniques

Many techniques have been successfully devised for variance reduction in general Monte Carlo settings but are less well known in the rendering community since their uses are not widespread, or are difficult and yet to be identified. Control variates reduces variance by utilising a function

$h(X)$ that is correlated with $f(X)$ that is easy to evaluate and with known expected value. Applying such techniques for rendering has been studied by for example in [SSSK04, FCH+06]. Applying techniques such as control variates and antithetic variables [Che81], where instead a negative correlation is constructed to reduce variance, to global illumination is non-trivial since correlated and easy to evaluate integrals are difficult to obtain. However, it is likely that many such techniques applied to problems in the physical and biological sciences still exist that can be of benefit to computer graphics and reduce variance in global illumination.

Conclusion This chapter has provided a background to Monte Carlo integration; one of the key theoretical concepts for solving the light transport problem. It has briefly covered the use of random variables and stochastic processes to solve high dimensional and complex integrals, and their advantages to the alternative of deterministic approaches. It has covered some important properties of Monte Carlo estimators such as convergence, bias and variance. In addition, techniques to combat variance such as importance sampling and its derivatives have been discussed and the need for good sampling patterns to generate well distributed random variables. Russian roulette allows the evaluation of infinite integrals, such as the rendering equation, and the application and use of these techniques in global illumination is the topic of the next Chapter.

Chapter 4

Paradigms for solving light transport

Contents

4.1 Monte Carlo path tracing **42**
4.1.1 Building on MC ray tracing 44
4.1.2 Path tracing 45
4.1.3 Sampling and path densities 49
4.1.4 Path termination 52
4.1.5 Particle tracing 52
4.1.6 Bi-Directional path tracing 54
4.1.7 Metropolis Light Transport 57
4.1.8 Trading accuracy for performance 59
4.2 Photon Mapping **60**
4.2.1 Photon tracing and storage 60
4.2.2 Radiance estimates 62
4.2.3 Rendering 63
4.2.4 Bias, variance and consistency 64
4.2.5 Limitations of classical photon mapping 66
4.2.6 Progressive Photon Mapping 67
4.3 Alternative Monte Carlo methods **72**
4.4 Extending unbiased Monte Carlo methods **73**
4.4.1 Cache-based importance sampling 74
4.4.2 Adaptive rendering 75
4.4.3 Sample reconstruction 76
4.4.4 Noise removal 77
4.4.5 Image based filtering 78
4.5 Conclusion **81**

Now that an theoretical understanding and an appreciation for the importance of probabilistic methods and Monte Carlo sampling in rendering has been established, this chapter will discuss in-depth a number of paradigms and practical algorithms for evaluating the rendering equation. In order to develop novel algorithms to solve light transport and to assess their impact in the current field of research, a number of popular and state-of-the-art techniques must be discussed. This chapter provides a focussed analysis of a handful of such techniques, including their advantages and drawbacks which will be built upon by the novel work forming the remainder of the thesis. Unbiased Monte Carlo path tracing methods and the popular photon mapping approach are two such successful techniques that are both widely used and receive a large amount of attention in the literature.

Quadrature-based integration Classic numerical quadrature methods such as interpolation or extrapolation of polynomials do not typically extend well to multi-dimensional problems. To maintain a given level of accuracy for such techniques, sampling must increase exponentially with the dimension of the integral. In Section 2.4.3 it was established that light transport is a problem of infinite dimensionality; light can be scattered between surfaces in scene indefinitely, until being completely absorbed. Thus solutions that rely on deterministic methods require truncation of the integral into lower dimensions, resulting in an incomplete solution. Alternatively, quantisation can be applied to produce a discrete representation of the true domain, again reducing the accuracy of the solution.

Finite element methods such as radiosity [GTGB84, CG85] quantise scene surfaces into patches, analytically computing the flux transfer between patches. A set of form factors are computed for each patch that describe the light transfer coefficients between two patches based on their visibility and orientations. Indirect illumination is accounted for by iteratively propagating flux received by a patch to all other patches, scaled by the precomputed form factors. Radiosity itself is thus not robust in high frequency integrals, and requires increasing numbers of patches to handle hard phenomena. The remainder of the techniques discussed in this chapter (and the thesis) are based on stochastic sampling of the integrand, which is used consistently across state of the art methods, as we will see.

4.1 Monte Carlo path tracing

The final part of Chapter 2 presented light transport as an integration problem in the form of Kajiya's rendering equation [Kaj86]. Chapter 3 then discussed the power of Monte Carlo integration for such problems, where analytical methods cannot handle the complexity and level of detail required to evaluate it accurately. Even MC methods such as the distributed ray tracing ideas of Cook et al.[CPC84] cannot directly handle the infinite dimensions of the rendering equation without truncation to lower dimensions, adding systematic error to the solution.

It was not until the introduction of Russian roulette to rendering by Arvo and Kirk [AK90] that an unbiased means of evaluating the infinite sum of terms was established, solving the full rendering equation without bias. Combining Russian roulette with stochastic ray tracing allows the path tracing algorithm of Kajiya [Kaj86] to handle the continuous high dimensional

domain that is presented by complex geometry, materials and lighting configurations required for realistic rendering. A substantial amount of work has been based around these three key theoretical concepts, increasing the capabilities of light transport simulation. This section will discuss in detail the fundamental theoretical and practical aspects of stochastic, path-based global illumination and the state-of-the-art extensions that have emerged.

Recall (from Section 2.4.3) the rendering equation expressed as an integral over surface area:

$$L_r(x' \to x) = L_e(x' \to x) + \int_S L_r(x'' \to x') f_r(x'' \to x' \to x) G(x'' \leftrightarrow x') V(x'' \to x') \mathrm{d}A(x'') \quad (4.1)$$

The reflected radiance term $L_r(x' \to x)$ is modified by the terms of the BRDF and the geometry terms necessary to account for the respective orientation of x and x' and the visibility between them. These terms are more conveniently referred to as the *throughput* at x' represented as $T(x \to x' \to x'')$, or $T(x', \omega, \omega')$ when dealing with solid angle measures:

$$\begin{aligned} T(x', \omega, \omega') &= f_r(x', \omega, \omega') \cos(\mathbf{n}_{x'}, \omega') \\ &= T(x \to x' \to x'') \\ &= f_r(x'' \to x' \to x) G(x'' \leftrightarrow x') V(x'' \to x') \end{aligned}$$

To obtain an equation for the radiance incident at x' scattered from x'', the $L_r(x'' \to x')$ term in the integral can be substituted by the right hand side of the equation, known as Neumann series expansion. Intuitively, the radiance arriving at x' must have been either emitted at x'' or scattered towards x'' from additional surfaces in the scene. Thus for an infinite length path $\bar{x}_\infty \in \Omega_\infty$ where $\bar{x}_\infty = x_0, x_1, ..., x_\infty$, Equation 4.1 can be rewritten as the sum of the radiance from its constituent sub-paths with length $k \leq \infty$:

$$\begin{aligned} L_r(x_1 \to x_0) = \sum_{k=1}^{\infty} \int_S \int_S ... \int_S L_e(x_k \to x_{k-1}) \\ \times T(x_k \to x_{k-1} \to x_{k-2}) ... T(x_2 \to x_1 \to x_0) \mathrm{d}A_k ... \mathrm{d}A_0 \end{aligned} \quad (4.2)$$

Known as the *path integral formulation* of light transport, its evaluation requires a mechanism able to construct and evaluate the throughput of paths for any finite length k. The idea of throughput at a vertex can be extended to representing the throughput of the path $\bar{x}_k = x_0, ..., x_k$ from its origin x_0, to its endpoint x_k:

$$\begin{aligned} T(\bar{x}_k) &= \prod_{i=1}^{k} T(x_{i-1} \to x_i \to x_{i+1}) \\ &= \prod_{i=1}^{k} f_r(x_{i-1} \to x_i \to x_{i+1}) G(x_{i-1} \leftrightarrow x_i) V(x_{i-1} \to x_i) \end{aligned}$$

Combining the path formulation of light transport with the measurement equation (Equation 2.7) provides a more intuitive equation of light transport as the basis of path based methods, as an integral of surface areas $A_{x_0}, ..., A_{x_k}$:

$$\begin{aligned} M_p &= \int \int W_e(x_0 \to x_1) L(x_1 \to x_0) G(x_0 \leftrightarrow x_1) \mathrm{d}A_{x_0} \mathrm{d}A_{x_1} \\ &= \sum_{k=0} \int_{A_0} ... \int_{A_k} W_e(x_0 \to x_1) G(x_0 \leftrightarrow x_1) \\ &\qquad \times T(\bar{x}_k) L_e(x_{k+1} \to x_k) \mathrm{d}A_{x_0} ... \mathrm{d}A_{x_k} \end{aligned}$$

Thus to solve global illumination, a method is required that can evaluate paths of any length k, connecting the camera and light sources to obtain W_e and L_e respectively, and evaluate the radiance measurement M_p for each pixel. However, since the incoming radiance and outgoing radiance at a point are in equilibrium, $T(x_1),...,T(x_k)$ are symmetric. A useful property results; valid paths can be formed by connecting the light source and the camera, starting at any vertex along the path. This provides substantial flexibility in algorithm design and opportunities to improve efficiency as shown in the remainder of this chapter.

4.1.1 Building on MC ray tracing

Path tracing is an extension to classic Whitted ray tracing and Cook's distributed ray tracing, that extending paths beyond the first non-specular vertex to account for illumination arriving indirectly.

Starting with a location on the camera sensor x_0, a ray through the lens is generated and intersected with the scene objects to find a point x_1 on a visible surface. If this surface is specular, as with Whitted's model, the path is continued until a non-specular vertex or light source is found (or until the ray leaves the scene). In path tracing, the path is continued to compute the incident illumination at indirectly visible points, accounting for the indirect illumination arriving at the first non-specular vertex x_d.

In Monte Carlo ray tracing, an estimate of the direct illumination incident at any surface position x' in the scene can be obtained by tracing additional rays from x' and recording their radiance contributions, evaluating $L_r(x',\omega)$. If N rays are traced stochastically over the hemisphere at x' with directions ω_i' distributed according to some PDF $p(\omega')$, a Monte Carlo estimate for $L(x',\omega)$ is obtained:

$$\begin{aligned} L_{direct}(x',\omega) &= L_e(x',\omega) + \int_{\Omega_{x'}} L_e(r(x',\omega_i'),-\omega_i') f_r(x',\omega,\omega_i') \cos(\mathbf{n}_{x'},\omega_i') \mathrm{d}\omega_i' \\ &\approx \\ \langle L_{direct}(x',\omega)\rangle_N &= L_e(x',\omega) + \frac{1}{N}\sum_{i=1}^{N} \frac{L_e(r(x',\omega_i'),-\omega_i') f_r(x',\omega_i',\omega) \cos(\mathbf{n}_{x'},\omega_i')}{p(\omega_i')} \end{aligned}$$

where $\langle L_{direct}(x',\omega)\rangle_N$ is an estimate of the direct lighting at x' obtained by evaluating N samples. Assume $x' = x_d$, the first diffuse hit point of the path. Evaluating $\langle L_{direct}(x',\omega)\rangle_N$ provides an estimate for the direct lighting at x' scattered towards x. In order to compute global illumination, the radiance scattered along $r(x',\omega_i')$ towards x' must also be included; the indirect term $L_r(r(x',\omega_i'),-\omega_i')$. Substituting $L_e(r(x',\omega_i'),-\omega_i')$ inside the sum with the total radiance $L(r(x',\omega_i'),-\omega_i')$ along $-\omega_i'$:

$$L(r(x',\omega_i'),-\omega_i') = L_e(r(x',\omega_i'),-\omega_i') + L_r(r(x',\omega_i'),-\omega_i')$$

allows both the direct and indirect lighting to be accounted for. Evaluating $L(r(x',\omega_i'),-\omega_i')$ can be achieved by finding the closest intersection point of the ray $r(x',\omega_i')$ and evaluating the direct lighting once again using Equation 4.3. Repeating this recursively produces a ray tree with a branch factor N connecting the camera sensor with the light source(s) via surfaces in the scene. An arbitrary N can be chosen at each vertex, to provide a radiance estimate for that

particular dimension of the integral. For large N the number of rays can become prohibitively costly. The radiance carried by each path is diminished by $T(x \to x' \to x'')$ at each vertex due to energy conservation, thus longer paths also have less impact on the final image than shorter ones. As a result N is usually kept at one for each vertex, producing cheaper computational costs per path, which also enables the scene space to be explored more thoroughly permitted by the increased number of independent paths.

The foundations for MC path tracing and the algorithms in the following section are reliant on stochastic path generation, thus it is important to understand their construction and characteristics.

4.1.2 Path tracing

The path tracing algorithm treats each pixel of the image as a separate integral, and a Monte Carlo estimator for each provides an expected value for the measured radiance through that pixel on the camera sensor. Thus paths are constructed for each pixel independently, exploring the scene and gathering radiance for a single pixel via numerous independent random walks. Supposing a path is constructed from some point on the camera sensor x_0 in the direction of the lens ω_0' to a visible point $r(x_0, \omega_0') = x_i$ in the scene. Extending from x_i requires a number of steps to be performed:

1. Obtain a new outgoing direction ω_i' over the hemisphere at x_i,
2. Update the path throughput $T(\bar{x}_i)$,
3. Check for path termination by applying Russian roulette to $T(\bar{x}_i)$,
    a. If the path is terminated, return.
    b. Otherwise trace a ray $r(x_i, \omega_i')$ to obtain x_{i+1},
4. Scale $T(\bar{x}_i)$ to account for Russian roulette,
5. Set $x_i = x_{i+1}$ and $\omega = -\omega'$ and repeat from 1.

These steps are repeated until the path is terminated either by Russian roulette or upon the ray leaving the scene. At each vertex x_i, direct lighting can be computed using Equation 4.3, and scaled by the paths throughput $T(x_0 \to x_i)$ obtaining an unbiased estimate for the measured radiance through p. This procedure is outlined in Algorithm 1.

The inclusion of indirect lighting significantly complicates the evaluation of each pixel integral, so variance reduction techniques become increasingly important when dealing with the complex light transport that can occur when handling multiple-bounce illumination. At each step, a number of the variance reduction techniques discussed in Chapter 3 can be applied to drastically improve the efficiency of the path tracing algorithm. Understanding the principles behind these techniques is crucial to identify their drawbacks and develop novel variance reduction techniques later in the thesis.

Algorithm 1 Path extension for global illumination

1: $\omega_i' \leftarrow$ sample hemisphere over x_i
2: $T(\bar{x}_i) = T(\bar{x}_{i-1}) \times T(x_i, \omega_i, \omega_i')$ ▷ Update path throughput
3: $\xi \leftarrow$ random variable in $(0,1]^{\mathbb{R}}$
4: $q = min\{q, Lum(T(\bar{x}_i))\}$
5: **if** $\xi > q$ **then** ▷ Apply Russian roulette
6: terminate path $\bar{x}_i$
7: **else**
8: $T(\bar{x}_i) \leftarrow T(\bar{x}_i) \,/\, q$ ▷ where q is the termination threshold
9: $x_{i+1} \leftarrow r(x_i, \omega')$ ▷ Trace ray to find x_{i+1}
10: $\omega_i \leftarrow -\omega_i'$

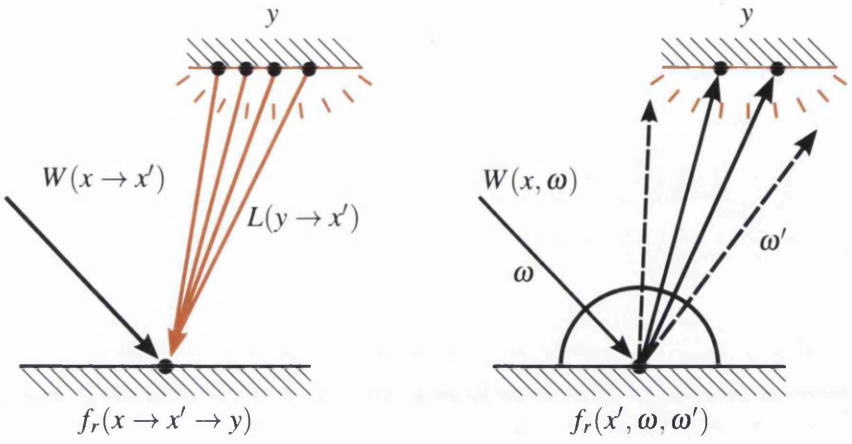
Explicit and implicit connections So far, direct lighting has been computed by a ray $r(x_i, \omega')$ sampled over the hemisphere of x_i using $p(\omega')$ hitting a light source during path extension. These are known as implicit connections, since they are a result of the path extension mechanism.

For low frequency BRDFs the distribution of $p(\omega')$ is wide, and as a result the probability of hitting a light source over x_i is small, since light sources typically do not take up a large portion of the hemisphere over a surface. A better approach is to keep track of the emissive surfaces in the scene (the light sources) and sample from their surfaces stochastically. The sampled location y on the light source can then be joined with the camera path vertex using a shadow ray, $r(x_i, x_i \rightarrow y)$, to account for visibility. The BRDF can be evaluated using $f_r(x_i, \omega, x_i \rightarrow y)$ to calculate the radiance contribution at x_i.

These explicit connections are linked to the idea of *next event estimation*, whereby the next event in the random walk (the path) is obtained by sampling from a distribution other than those defined over the current state, which propagate the dual quantity carried by x_i. In the case of ray tracing and path tracing, which carry importance, explicit connections sample from radiance distributions over the light sources to obtain $x_{i+1} = y$.

However, explicit connections are not always applicable. For diffuse BRDFs, which are low frequency and where $f_r(x_i, \omega, \omega') \neq 0$ for most ω', they can vastly reduce variance. For higher frequency BRDFs that are zero across large regions of the hemisphere they cannot be applied efficiently since it is highly likely that any ω' chosen from a distribution other than one from the BRDF will yield zero radiance. As the BRDF becomes more restrictive, the problem is exacerbated and so an alternative approach must be taken to deal with singularities that follow Dirac distributions. Perfect specular reflection and refraction are common examples of Dirac distributions in rendering, and as a result implicit lighting is the only method for accumulating radiance scattered by such surfaces.

Singularities and Dirac distributions Singularities such as specular interactions are commonly encountered in rendering. Though they are not physically valid, conveniences such as the infinitesimal surface area of point light sources and the emission direction of distant light sources are also represented by Dirac distributions and cause problems for light source



(a) Explicit connections

(b) Implicit connections

Figure 4.1: *Direct lighting at a point in the scene can be computed explicitly, by sampling a point y on the light source (a). Alternatively, sampling from the BRDF at the vertex* $f_r(x \to x' \to y)$ *produces implicit connections.*

sampling.

Although they help to reduce variance for a particular area of the integrand, since they introduce a deterministic step to an otherwise stochastic path construction process, singularities can also cause problems; like inhibiting the ability to obtain radiance or importance measurements explicitly via shadow rays discussed above. This is the root of many problems for light transport algorithms when combined with the variance introduced by stochastic path tracing through lower frequency BRDFs, that require implicit connections to sample the radiance or importance flowing through them.

The sequences of specular and diffuse vertices in light transport paths causes problems for many methods, and so sampling all light paths with low variance is still an open problem, and one that will be discussed in more detail and motivates the novel work proposed in Chapter 5.

4.1.2.1 Path correlation

The path construction and radiance evaluation described this far is a naïve but theoretically valid one: a path from x_0 is constructed by repeatedly extending the path to x_k until hitting a light source (denoted y), and evaluating the measurement function at x_k.

This approach is highly inefficient, since path construction comes at considerable cost due to the ray tracing function and so intuitively, improved efficiency can come about by increasing the ratio of radiance measurements to traced rays.

A better approach is therefore to estimate the incident direct lighting at each vertex, thereby creating a correlated series of radiance contributions that reuse the same path vertices. Obtaining multiple correlated contributions from a path of k vertices, requires a maximum of $2k$ rays,

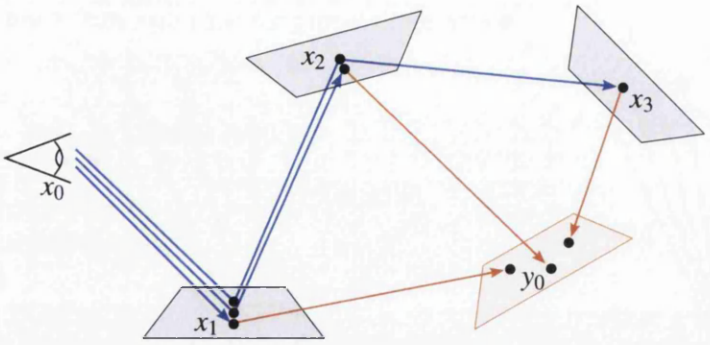


Figure 4.2: *Tracing paths naïvely as illustrated here is computationally expensive. This can be reduced by making use of path correlation across the integrand, reusing vertices of longer paths to compute radiance estimates for multiple shorter paths.*

assuming every vertex is non-specular. To obtain the same contributions, a set of naïve paths require $\frac{k(k+1)}{2}$ rays, since each radiance contribution comes from a new random walk which must be generated individually from x_0, making it far less efficient (Figure 4.2).

In conjunction with explicit light source sampling, this potentially enables a radiance measurement to be obtained at each non-specular vertex using next event estimation, whilst reusing the current state of the random walk to produce a correlated radiance contribution stochastically.

Utilising this correlated approach requires careful handling of samples to ensure each path is weighted correctly. For diffuse surfaces, direct lighting computed explicitly by sampling y_0 will commonly provide most of the radiance contributions. However in the interest of reducing variance and handling specular surfaces where narrow BRDFs inhibit explicit lighting, the inclusion of implicit connections is still necessary. Assume that a contribution from $L_e(x_k \to x_{k-1})$ of a path $x_0, ..., x_k$ is accounted for implicitly at x_k by path extension. Using an explicit connection, the same path space may be accounted for as direct lighting at x_{k-1} by sampling y_0 as a light source (see Figure 4.3). To simplify this, the integrand can be separated using the notion of path space such that a path of length k only evaluates a single sample from each set $\Omega_0, \Omega_1, ..., \Omega_k$, obtaining a single contribution for each path length in $(0,k)$. Thus if the vertex at x_{k-1} has sampled a light source explicitly, $L_e(x_k \to x_{k-1})$ should not be added since the radiance of a path in Ω_k has already been accounted for.

Some important nomeclature Generally, a Monte Carlo sample in the context of rendering algorithms refers to the correlated set of paths and their combined radiance measurement arriving at the camera x_0 along the primary ray:

$$L(x_0, \omega_0) = \sum_{i=0}^{k} T(\bar{x}_i) L_e(x_i, \omega_i) L_r(x_{i+1}, \omega_i)$$

since they are commonly generated consecutively using a set of random variables and the path extension mechanism described above. A *sample contribution* or the *sample radiance* therefore refers to the total radiance arriving at the camera sensor from this correlated set of paths. This

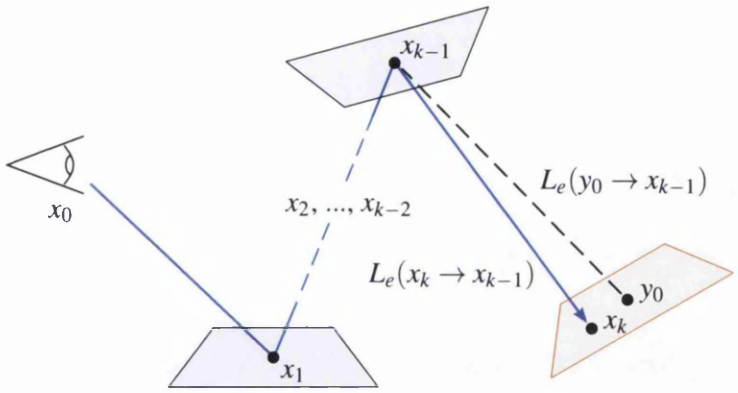


Figure 4.3: *Example of a path $\bar{x}_k = x_0, x_1, \ldots, x_k$ constructed from the camera. When evaluating both implicit (from $x_{k-1} \rightarrow x_k$) and explicit ($x_{k-1} \rightarrow y_0$) connections, care must be taken to ensure each path length is only accounted for once. Direct lighting for $x_0, x_1, \ldots, x_{k-2}$ is omitted for clarity.*

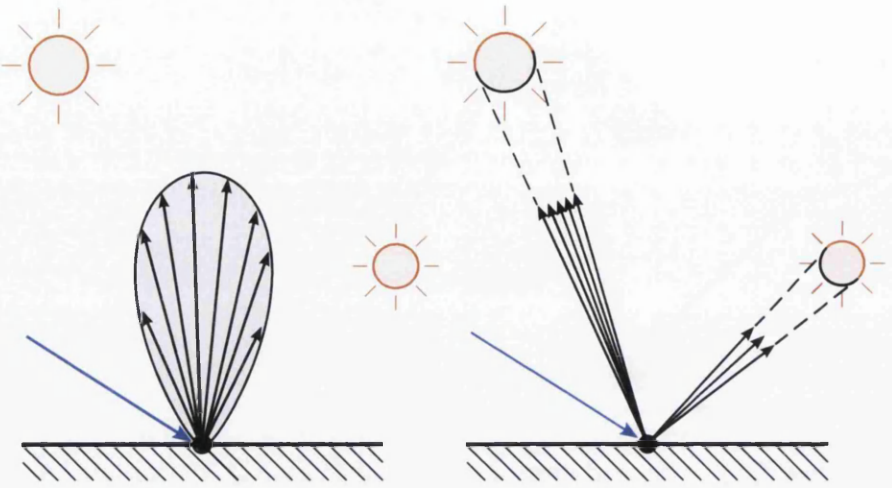
sample radiance is comprised of the accumulation of radiance at each vertex x_i for all $0 \leq i \leq k$ each of which represents a single path in Ω_i connecting x_i with a point on the light source. These are described as the *vertex contributions* of the sample. This is an important concept for the techniques introduced in this thesis based on the isolation and utilisation of path vertex data, and will be discussed in further detail in later chapters.

4.1.3 Sampling and path densities

Ensuring good sample generation is an important aspect of MC based rendering, driving every stochastic step of the process. Producing high quality sampling patterns for pixel sub-sampling reduces geometry and texture aliasing effectively, but including indirect lighting emphasises the need for good stratification across higher dimensions. Although each pixel is treated as an independent integral, sample distributions can take into account the local pixel neighbourhood in order to prevent excessive correlation, and by doing so can reduce perceptual artefacts and noise. In addition, due to the diverse nature of BRDFs, an initially well stratified distribution can easily breakdown as it becomes warped by surface interactions, minimising its effectiveness to explore the scene space. As a result, the importance sampling and variance reduction techniques discussed in Section 3.3 can have a high impact on the final visual quality.

4.1.3.1 Importance sampling

Importance sampling is one of the widest used variance reduction techniques in global illumination, as there is a significant amount of information about the shape of the integral available prior to obtaining radiance calculations. Both sensor importance in the form of BRDF distributions, and light source positions and flux are known at any given path vertex. Path construction relies on the ability to generate an outgoing direction $x_i \rightarrow x_{i+1}$ such that the radiance received along it is maximised, aligned with the desire to concentrate samples where the integral func-



(a) BRDF importance sampling (b) Light source importance sampling

Figure 4.4: *Illustration of importance sampling on a diffuse surface, using (a) the BRDF distribution and (b) using a distribution over the surface area of the light sources. For wide BRDFs, light source sampling is far more effective than relying implicitly on the BRDF distribution to obtain radiance from light sources.*

tion $f(X)$ and hence $L(x_{i+1} \to x_i)T(\bar{x}_i)$ is largest. This provides two possibilities in determining suitable ω':

1. Assume $L(x_{i+1} \to x_i)$ is constant for all possible x_{i+1} and sample using $T(x_{i-1} \to x_i \to x_{i+1})$ to obtain x_{i+1}.
2. Estimate $L(x_{i+1} \to x_i)$ by choosing x_{i+1} directly from the surface of light sources.

In the case of path extension, no knowledge of the indirect radiance distribution over the hemisphere at x_i is known, and hence $T(x_{i-1} \to x_i \to x_{i+1})$ must be relied upon as an approximation to $f(X)$ to reduce variance. Recall that path throughput can be broken down into the BRDF $f_r(x_{i-1} \to x_i \to x_{i+1})$, geometry term $G(x_{i+1} \leftrightarrow x_i)$ and visibility $V(x_{i+1} \leftrightarrow x_i)$. For Lambertian surfaces where the BRDF is uniform, sampling can be performed according to the geometry term, concentrating samples around the surface normal of x_i.

Given that for the last vertex of a path x_i the incident direction $x_{i-1} \to x_i$ is known, using a PDF that follows the distribution of the BRDF at x_i to sample ω' will maximise the throughput of the path $T(\bar{x}_{i+1})$ for radiance calculations at x_{i+1}.

When computing direct lighting at a non-specular vertex, the positions and emitted flux of each light source is available and hence only the visibility term $V(x_{i-1} \to x_i)$ remains unknown. Ideally, ω' would be sampled according to the product of $L(x_{i+1} \to x_i)$ and $f_r(x_{i-1} \to x_i \to x_{i+1})G(x_{i+1} \leftrightarrow x_i)$, however this is difficult to achieve in practice.

Using $T(x_{i-1} \to x_i \to x_{i+1})$ to approximate $f(X)$ as used during path extension can suffice when handling highly-specular BRDFs, since a narrow BRDF produces larger regions of the

integrand where $T(x_{i-1} \rightarrow x_i \rightarrow x_{i+1})$ will be zero. In the case of diffuse surfaces, sampling according to a PDF based on the position and flux of the light sources is often better, since diffuse BRDFs are commonly non-zero everywhere. Explicit direct lighting follows this logic and is a form of importance sampling according to the distribution of L over all light sources. Ideally, a combination of these approaches would reduce variance further, despite the inability to draw samples from $L(x_{i+1} \rightarrow x_i)T(x_{i-1} \rightarrow x_i \rightarrow x_{i+1})$ directly.

Sampling using all information available from multiple distributions can be achieved by the use of multiple importance sampling (MIS, recall Section 3.3). Using MIS, samples can be drawn independently from each distribution and weighted based on the probability of choosing each sample from the opposing distribution. The disadvantage with this approach is that the full contribution of each sample must be evaluated, requiring rays to be traced from each distribution handled by MIS. However the reduction in variance can be significant, especially in the presence of glossy surfaces or small light sources. In practice, the expensive evaluation of $V(x_{i-1} \rightarrow x_i)$ for each sample can be delayed until after $L(x_{i+1} \rightarrow x_i)f_r(x_{i-1} \rightarrow x_i \rightarrow x_{i+1})G(x_{i+1} \leftrightarrow x_i)$ is determined to be non-zero, which is cheaper to compute.

Singularities in the BRDF caused by specular reflection and refraction at x_i implicitly require importance sampling when sampling from the BRDF since there is only one x_{i+1} for which $f_r(x_{i-1} \rightarrow x_i \rightarrow x_{i+1})$ is non-zero and finite, which is required for a valid PDF $p(X)$ to exist. Additionally, MIS provides no benefit since there is only a single distribution from which samples can be drawn.

Path densities Each radiance contribution $L(x_{i+1} \rightarrow x_i)\ T(x_{i-1} \rightarrow x_i \rightarrow x_{i+1})$ is dependent on the random variables used to extend the path from x_i to x_{i+1}. The application of local importance sampling (for BRDFs or direct lighting) means that an independent PDF is used to extend the path at each vertex. The probability of choosing x_{i+1} given x_i is denoted $p(x_{i+1}|x_i)$ abbreviated to $p(x_i)$. For a path of length k terminated by Russian roulette, the unbiased radiance measurement for this path R_k must be weighted by $p(x_i)$ at each vertex $0 \leq i < k$:

$$R_k = \frac{T(\bar{x}_k)}{p(\bar{x}_k)} = \prod_{i=0}^{k} \frac{T(x_i)}{p(x_i)} \tag{4.3}$$

$p(\bar{x}_k)$ is known as the *path density* of $\bar{x}_k = x_0, ..., x_k$. To be a valid PDF, it must hold that $p(x) \leq 1$ for all x. Unless p is deterministic (ie: undergoes a perfect specular interaction) as k increases $p(\bar{x}_k)$ is decreased, resulting in longer paths becoming increasingly more difficult to sample stochastically. The notion of path density can be generalised to describe the *partial density* for a subset of the path between any two vertices x_a and x_b:

$$p(x_a \rightarrow x_b) = p(x_a)p(x_{a+1}|x_a)...p(x_b|x_{b-1}) = \prod_{i=a}^{b} p(x_i|x_{i-1}) \tag{4.4}$$

High variance often occurs when the PDF is a poor estimate of $L(x_{i+1} \rightarrow x_i)T(x_{i-1}, x_i, x_{i+1})$ at x_i. When using locally greedy importance sampling of the BRDF $p(x_i|x_{i-1}) \propto T(x_{i-1} \rightarrow x_i \rightarrow x_{i+1})$ ignoring the effect of $L(x_{i+1} \rightarrow x_i)$. When paths with low density are generated for which

$L(x_{i+1} \to x_i)$ is large, because $p(\bar{x}_k) \ll 1$ it is also the case that $R_i \gg L(x_{i+1} \to x_i)$, increasing the pixel estimators variance $V[\langle I_p \rangle]$. This high variance was discussed in Section 3.2 and is responsible for the spiked noise that is commonly associated with Monte Carlo methods. In rendering, this appears in the form of bright speckles over individual pixels in the image.

4.1.4 Path termination

Russian roulette makes it feasible to evaluate an unbiased approximation to the rendering equation, in addition to providing improvements in efficiency when used correctly. Terminating paths deterministically after a fixed number of bounces is common in interactive renderers, as it provides more predictable performance characteristics, however it also adds bias to the solution, ignoring radiance measurements at longer path lengths. This becomes noticable in highly occluded scenes and in specular and glossy environments where longer paths carry larger radiance contributions.

In its simplest form, Russian roulette can be applied by drawing a uniform random variable ξ and comparing it with a constant q, both in $[0, 1)^{\mathbb{R}}$ to probabilistically decide path termination at a vertex x_i. This provides an equal probability q of terminating a path at every vertex, regardless of its density or the importance it carries. Short paths are therefore favoured by Russian roulette and, since they generally carry higher importance (hence the potential for $f(X)$ to be large) this is desirable. However, due to the diversity of BRDFs commonly used in physically based scenes, especially in the presence of specular surfaces, this relationship between importance and path length no longer holds. An alternative approach is to use a different value for q at each path vertex that is proportional to the path importance $W_e(x_0 \to x_1)$, propagated along $\bar{x}_{i+1}$, equal to the path throughput $T(\bar{x}_{i+1})$. This dictates the proportion of radiance arriving at x_i that will be propagated back towards the camera, and thus is a good measure of the paths potential to contribution to the image, without knowledge of the incident radiance. When deciding path termination at x_i the path throughput is updated as:

$$T(\bar{x}_{i+1}) = \begin{cases} T(\bar{x}_{i+1}) / t & \text{if } \xi < t \\ 0 & \text{otherwise} \end{cases}$$

where t is a scalar representation of $T(\bar{x}_{i+1})$ such as its luminance or maximum value across each channel that represents the colour spectrum (typically an RGB triplet). This ensures that each path is weighted to account for the radiance potentially lost by those paths that are terminated, ensuring over time that statistically, the expected value of the radiance estimator is equivalent.

4.1.5 Particle tracing

Also known as light tracing, this algorithm was introduced by Dutré [DLW93] and follows the same principles as path tracing, but reverses the direction of the path construction by starting from the light sources. Instead of propagating importance, light tracing simulates the propagation of radiance through the scene, more closely modelling photons in the real world. This allows it to better account for high intensity light paths such as caustics, that are likely to have far higher path densities when sampled from the light source, as opposed to the camera. Due

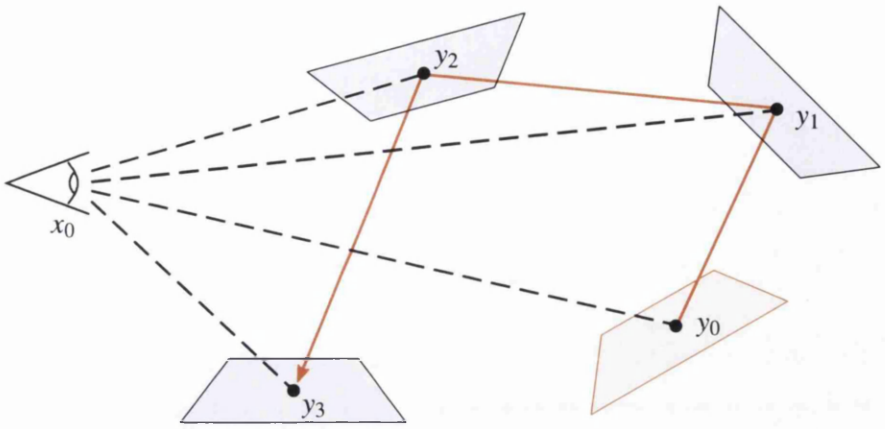


Figure 4.5: *Light tracing is the reverse of path tracing, extending paths from the light sources in the scene and explicitly connecting them to the camera.*

to the reciprocity of BRDFs, and the symmetry of occlusion and geometric terms in the rendering equation, whether tracing from the light or the camera the same integral is solved, and the same path construction framework can be employed. Path tracing generates a number of samples through each pixel of the image, corresponding to a separate estimator for each individual pixel. Intuitively it makes sense to devote an equal number of samples to each pixel, as they all share an equal portion of the image space [1]. In light tracing, and indeed any method that propagates radiance starting from the light sources alone, the whole domain is treated as a single integral, increasing the significance of good sampling strategies. First, for each particle a position on a light source must be selected. Commonly, an importance sampling strategy is used that selects a light source based on the total flux distribution over all light sources. Thus emitters of higher power (eg: large weak emitters, or small bright ones) will emit proportionally more particles and hence contribute more to the final image. Using a uniform sampling strategy would require re-weighting each particle according to the light sources power, increasing the variance between emitted particles, and hence variance in the final image.

A point y_0 on the light is then selected uniformly over the surface. Finally, an outgoing ray direction is chosen assuming a Lambertian BRDF at y_0, or with respect to IES lighting data if applicable. From y_0 the ray is traced to obtain the point y_1 lit directly by this photon.

Path extension is performed as before from y_1 (recall Algorithm 1) by sampling the BRDF, applying Russian roulette and updating the path throughput until termination, forming a light path $y_0 \ldots y_l$. Equivalent to the explicit direct lighting computations used in path tracing, explicit connections from each light tracing vertex can be made to the camera lens, in order to determine the pixel that should receive its contribution. In a pinhole camera model, since the lens is represented by a single point, the correct pixel is simply found by calculating the intersection point of a ray $r(y_l, y_l \rightarrow x_0)$ with the camera sensor. Just as path tracing must rely on implicit

[1] This is not necessarily optimal, since the goal is to minimise the variance of the image as a whole, which is usually not equally distributed between pixels. Section 4.4 discusses this in more detail along with alternative approaches to pixel sampling

path extensions to accumulate radiance incident on specular surfaces from $x_k \rightarrow y_0$, light tracing requires implicit connections to connect specular vertices carrying reflected radiance to the camera sensor, $y_l \rightarrow x_0$. In a pinhole camera model this is impossible, since both the lens and the specular BRDF at the vertex are Dirac distributions. Even assuming a finite aperture model, the lens is not modelled geometrically and so implicit connections are still not feasible.

Though the algorithms of path tracing and light tracing are theoretically equivalent, in practice their performance can be significantly different. A major shortcoming of the light tracing algorithm is its inability to render reflections and refractions seen through specular objects. Scenes dominated by caustic lighting are handled well by light tracing, however caustics are only produced in the presence of highly reflective or refractive surfaces. Additionally, light tracing paths can focus on regions that have low or no importance to the image, where even the high radiance they carry will not contribute to the image, wasting computation. The same is true for path tracing with respect to low radiance regions, however it is generally more efficient due to the diffuse characteristics of light transport and the viewpoints used in the majority of scenes. This is discussed in more detail in Chapter 5 and is one of the principles behind our hybrid rendering approach.

4.1.6 Bi-Directional path tracing

The path tracing and light tracing algorithms trace a set of correlated paths, connecting each vertex to the light source or camera sensor respectively. In the case of path tracing, the radiance contribution in the path space of Ω_k is created by joining the path $\bar{x}_{k-1}$ with a light source, producing a path of the form:

$$x_0, x_1, \ldots x_{k-1}, y_0$$

where y_0 is a point on a light source obtained explicitly, or

$$x_0, x_1, \ldots x_k$$

where $x_k = y_0$ in the case of implicit connections, since x_k is generated as an extension of x_{k-1}. Light tracing paths are formed equivalently but by the extension of y_0 resulting in y_0, y_1, ... y_{l-1}, x_0 or in the implicit case y_0, y_1, ... y_l where $y_l = x_0$.

Under more difficult lighting conditions this can perform poorly, since connecting x_0 and y_0 by extending only one path is not always effective, and can result in the propagation of importance to regions with low radiance (and vice versa in the case of light tracing), without considering the result of the overall measurement equation which it is sampling.

Developed independently by Veach and Guibas [VG94] and Lafortune and Willems [LW93], *bidirectional path tracing* (BDPT) samples and extends paths from both the camera and a light source, connecting them explicitly, thereby sampling proportional to both radiance and importance. Given a camera path $\bar{x}_k$ of length k and a light path $\bar{y}_l$ of length l, their endpoints can be connected constructing a complete path of length $k+l+1$, between camera and light source:

$$x_0, x_1, \ldots x_k, y_l, \ldots, y_1, y_0$$

This attempts to maximise the throughput of the complete path and thus the measurement function that is being integrated. Taking the path integral formulation for the radiance measurement

at pixel p, this can be expressed as:

$$\begin{aligned} M_p &= W_e(x_0 \to x_1)\prod_{i=1}^{t} f_r(x_{i-1} \to x_i \to x_{i+1})G(x_{i-1} \leftrightarrow x_i)V(x_{i-1} \leftrightarrow x_i) \\ &\quad \times L_e(y_0 \to y_1)\prod_{i=1}^{s} f_r(y_{i-1} \to y_i \to y_{i+1})G(y_{i-1} \leftrightarrow y_i)V(y_{i-1} \leftrightarrow y_i) \end{aligned}$$

Given two independently generated paths $\bar{x}_t$ and $\bar{y}_s$, a bidirectional sample $\bar{z}_k = z_0, ..., z_k$ can be constructed using sub paths of $\bar{x}_t$ and $\bar{y}_s$, where $k = s+t$ and $0 \leq s \leq k$ and $0 \leq t \leq k$. Path tracing and light tracing can be thought of as special cases of this approach with $s \leq 1$ and $t = 1$ respectively [2]. By varying s and t while keeping $s+t$ constant, paths in Ω_k of equal length can be evaluated. A path of length n (with $n+1$ vertices) can be sampled in $n+1$ different ways, thus producing $n+1$ path densities, depending on where intermediate vertices are sampled from the light path or camera path. For a bidirectional path $\bar{z}_{s,t}$ given s and t the full PDF is:

$$p(\bar{z}_{s,t}) = p(x_0)...p(x_s|x_{s-1})p(y_t|y_{t-1})...p(y_0)$$

However, identical paths could be generated with multiple different path densities using the same PDFs:

$$\begin{aligned} p(\bar{z}_{s+1,t-1}) &= p(x_0)...p(x_s|x_{s-1})p(y_t|x_s)...p(y_0) \\ p(\bar{z}_{s+2,t-2}) &= p(x_0)...p(x_s|x_{s-1})p(y_t|x_s)p(y_{t-1}|y_t)...p(y_0) \\ &\ldots \\ p(\bar{z}_{s+t,0}) &= p(x_0)...p(x_s|y_t)p(y_t|y_{t-1})...p(y_1|y_2)p(y_0|y_1) \end{aligned} \tag{4.5}$$

requiring only the re-evaluation of the PDFs at each vertex in the reverse direction. Due to the reciprocity of BRDFs and symmetry of the rendering equation, the radiance R_p is unaffected by the sampling strategy and f remains constant regardless of the values of s and t used to obtain the path density. This provides an opportunity to reduce variance, by intelligently choosing a suitable PDF for the path. An estimator for the pixel integral can be obtained:

$$\langle I_p \rangle = \sum_{s\geq 0}\sum_{t\geq 0} w_{s,t}\frac{f(\bar{x}_{s,t})}{p_{s,t}(\bar{x}_{s,t})}$$

where $f(\bar{x}_{s,t})$ is the radiance measurement function M_p for the path $\bar{x}_{s,t}$ and $w_{s,t}$ is the weight given to a particular sampling strategy, to maintain an unbiased estimate. As for the case of explicit and implicit sampling, care must be taken to record only a single contribution for each path length up to k. Naïvely this can be done by summing the contributions of all paths of a given length n (with $n+1$ vertices), and dividing by the number of ways they can be constructed $n+1$, such that $w_{s,t} = \frac{1}{s+t+1}$. Averaging the estimators in this way means that together their variance is similarly a mean of the variance from each of the $n+1$ contributions, which is far from desirable [Vea97]. Multiple importance sampling can be applied to combine the strengths of each path construction strategy and minimise variance in the image, providing a set of good weights for each strategy $w_{s,t}$ of a path with length $s+t$.

[2] In the case of light tracing, the assumption is made that the lens is not represented geometrically and so implicit connections where $t = 0$ are impossible.

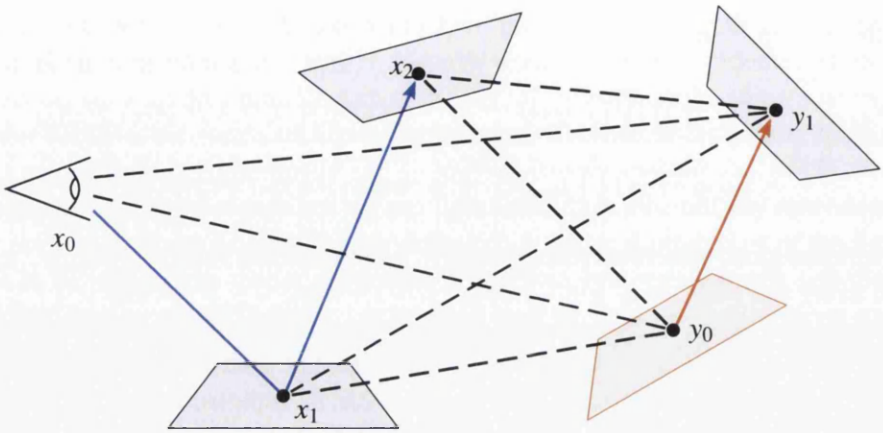


Figure 4.6: *A set of bidirectional paths can be constructed by explicitly joining the vertices of the camera and light paths.*

Simply connecting the endpoints of each path is inefficient, and in similar fashion to the unidirectional path and light tracing algorithms, a correlated set of paths from the camera to the light source(s) can be constructed by connecting intermediate vertices (Figure 4.6). In cases where light sources are highly enclosed (for example by a lamp shade) this is usually more efficient than relying on shadow rays from the camera path. The light path can first escape the enclosed space, and have a higher probability of linking to a camera path vertex in more open regions of the scene free from occluders. Although this requires additional rays to connect the vertices of each path combinatorially, in more difficult lighting scenarios it is often still advantageous. Veach [VG94] and Lafortune [LW93] both proposed a technique based on Russian roulette to skip some vertex connections if their contributions are minimal. By only connecting vertices with proportionally significant contributions, the number of rays required is reduced trading increased variance for improved efficiency.

The strength of BDPT comes from its ability to combine the advantages of path and light tracing, and using MIS to do so in a provably good way. Additionally, as both path tracing and light tracing are subsumed by bidirectional path tracing, it can make use of the explicit connections to light emitters and sensors to improve efficiency and performance. For light paths that are connected directly to the camera (where $s = 1$), a light path can contribute to multiple pixels by projecting vertices onto the camera sensor, instead of using the origin of the eye path generated for the bidirectional sample used for a particular pixel. Similarly, direct lighting (where $t = 1$) can be evaluated by choosing a point on the light source independent of the generated light path, instead using samples that are well stratified with respect to the surrounding pixels.

In some cases however, BDPT is restricted to evaluating paths identically to unidirectional methods, providing no additional advantage. In the presence of highly specular surfaces explicit connections cannot be made, restricting the number of strategies under which a path of given length can be constructed. Thus, the significant benefits provided by MIS are also eliminated.

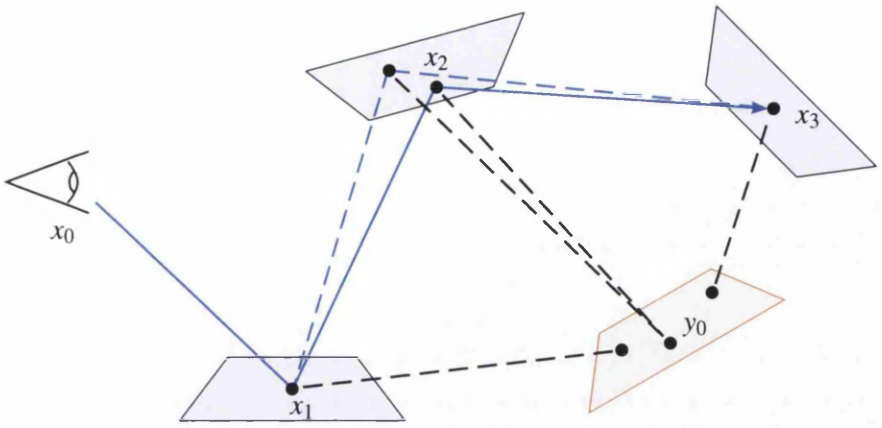


Figure 4.7: *Markov-chain MC algorithms like Metropolis light transport locally explore path space efficiently via path mutation, reusing sub-paths ($x_0 \rightarrow x_1$, $x_1 \rightarrow y_0$ and $x_3 \rightarrow y_0$).*

In highly occluded environments, many of the explicit ray connections used to construct the set of correlated paths can fail, decreasing the efficiency of BDPT. The camera and light paths are generated independently, so there is no guarantee that vertices on the two paths are mutually visible. Pajot et al.[PBPP11] propose a combinatorial approach to path connections replacing the one to one connections of camera and light paths with a many to many relationship, improving correlation and the use of each generated sub-path.

Since it is still a stochastic process, low density paths carrying high radiance still occur. Using locally greedy importance sampling strategies like BRDF sampling still present problems in highly occluded environments, where visibility is not factored into the sampling process. In cases where only a single evaluation strategy is available, BDPT does not provide any advantage over standard unidirectional techniques. Caustic lighting is a common situation where this occurs, especially when emanating from small light sources where the chances of implicit connections are reduced. To improve upon this, and in highly occluded environments, Markov-chain Monte Carlo methods can be used to explore regions of path space based on their measured importance, as opposed to a locally greedy estimate to the integral.

4.1.7 Metropolis Light Transport

Veach and Guibas [VG97] introduced the Metropolis-Hastings algorithm (discussed in Chapter 3) to light transport, providing an effective means to tackle difficult lighting scenarios. Here an introduction to the principle behind Metropolis Light Transport (MLT) is provided. For more in depth discussion see the Ph.D. thesis of Veach [Vea97] and the more recent work by Kelemen et al.[KSKAC02].

In Veach's original variant of Metropolis light transport, given an existing path (known as a state) $\mathbf{X}$ generated using MC sampling, a proposed path $\mathbf{X}'$ for the MC Markov-chain is generated by modifying (mutating) elements of the previous state $\mathbf{X}$. Evaluating this new state, and repeatedly applying mutations forms a chain of random walks through the integral, eventually generating a state with density proportional to f.

The tentative transition function $Tr(\mathbf{X} \rightarrow \mathbf{X}')$, is used to generate a proposed state $\mathbf{X}'$, representing the probability of obtaining $\mathbf{X}'$ from $\mathbf{X}$ under some mutation strategy.

This proposed sample $\mathbf{X}'$ is then accepted or rejected stochastically, using a pseudo-random variable ξ similar in operation to Russian roulette. The threshold for this is determined by an *acceptance probability* $a(\mathbf{X} \rightarrow \mathbf{X}')$ which is responsible for the behaviour and effectiveness of Metropolis sampling.

The acceptance probability must be designed such that an equilibrium state is reached for the implicit distribution obtained over the course of the Markov-chain, representing the ideal PDF over f. Thus, if $\mathbf{X}$ is already in equilibrium (where $p(\mathbf{X}) \propto f$) then $p(\mathbf{X}) = p(\mathbf{X}')$, otherwise mutations will tend away from the ideal PDF. This is described by the *detailed balance* rule:

$$f(\mathbf{X})\, Tr(\mathbf{X} \rightarrow \mathbf{X}')\, a(\mathbf{X} \rightarrow \mathbf{X}') = f(\mathbf{X}')\, Tr(\mathbf{X}' \rightarrow \mathbf{X})\, a(\mathbf{X}' \rightarrow \mathbf{X})$$

The efficiency of the Metropolis sampling estimator relies on the ability of the acceptance probability function to reach equilibrium with as few mutations as possible, which corresponds to maximising the acceptance probability:

$$a(\mathbf{X} \rightarrow \mathbf{X}') = min\left\{1, \frac{f(\mathbf{X}')Tr(\mathbf{X}' \rightarrow \mathbf{X})}{f(\mathbf{X})Tr(\mathbf{X} \rightarrow \mathbf{X}')}\right\}$$

If the sample is accepted, $\mathbf{X}$ is replaced by the new state $\mathbf{X}'$, otherwise $\mathbf{X}$ is kept as the basis for the next mutation. In either case, $\mathbf{X}$ is weighted using the acceptance probability $a(\mathbf{X} \rightarrow \mathbf{X}')$ and added to the estimator and the process continues, forming a chain of mutations where the distribution of $\mathbf{X}_i$ reaches an equilibrium distribution proportional to the ideal PDF; $f(\mathbf{X}) / \int_\Omega f(\mathbf{X})\mathrm{d}\Omega$.

When sampling path space stochastically, paths with low densities and high contributions are problematic, increasing pixel variance. Only a single sample is evaluated from these difficult regions before new samples are generated the estimator moves on to explore the remainder of the integrand. Using the path mutation approach of MLT, local path space regions with high contributions can be explored once found, contributing numerous samples despite their low densities. By reusing sub-paths of of previous samples that are unaffected by the mutation, exploration can be achieved whilst minimising additional ray tracing computation, and reducing the average sample cost.

Unlike pure MC path tracing, MLT does not evaluate individual integrals for each pixel but, using path mutation, distributes samples from a single estimator across the image. This can improve pixel correlation, since by mutating the path vertex that lies on the camera sensor, hard to find paths can contribute to multiple pixels via mutations. Radiance measurements of neighbouring pixels are often similar, so the correlation provided by path mutation can be beneficial. Variance for MLT estimators (especially for individual pixels) is non-trivial, and has been analysed in [APSS01].

Metropolis light transport has seen significant research attention. Kelemen et al.[KSKAC02] make a number of simplifications and improvements to MLT, sampling and mutating the random variables responsible for path generation within the n-dimensional hypercube over $[0, 1)^{\mathbb{R}}$. They also introduce the notion of large step mutations to improve ergodicity, in turn improving

stratification over the image. Sample stratification can be a problem with MLT, as acceptance probabilities traditionally sample pixels according to their radiance and ignore the distribution of samples with respect to individual pixels. In higher contrast scenes, darker pixels are often left undersampled and noisy while brighter regions may have adequately converged, such as directly visible light sources. These high contribution path remain the focus of the Metropolis sampler due to the tendency of the transition function.

Energy redistribution path tracing (ERPT) by Cline et al.[CTE05] combine standard MC path tracing with Markov-chain MLT, initialising independent MLT integrators for each pixel with short mutation chains, spreading sample contributions to nearby pixels in the image to improve correlation and variance. Shorter chains ensure the MLT integrator does not get stuck on high contribution paths for too long, avoiding bright splotches, and instead moving to new regions of the image plane to better explore the scene. A side effect of this is that the short chains cannot explore longer or more complex paths as thoroughly before moving on, increasing the overall convergence time for such scenes since there is a heavier reliant on stochastic sampling.

An alternative approach to improving stratification is to tailor the mutation and transition functions to suit the evaluation of multiple equally weighted integrals: the image pixels. Veach [Vea97] introduced two-stage MLT, which uses an initial low resolution and low sample density image to provide an estimate for the actual pixel integral, which is used as a target function for sample mutation and improve stratification. For difficult scenes this is impractical and may require a significant amount of computation to obtain an accurate estimate, which is then discarded upon starting the MLT sampler at full resolution. Hoberock and Hart [HH10] improve upon this with an iterative start up approach that increases the image size gradually, contributing all samples to the final image. Additionally, they propose a noise-aware variant of MLT which adjusts the target distribution based on a perceptual variance estimate, improving sample distribution and visible noise.

Recently, Jakob and Marschner [JM12] tackled the problem of mutations in highly specular paths, using the tight constraints of specular vertices to define a manifold in path space. Utilising the differential geometry at each surface, this manifold can be locally explored via path mutation, allowing it to fit in to existing MLT and ERPT algorithms providing significant improvements over previous mutation approaches.

4.1.8 Trading accuracy for performance

In some applications, the performance or noise characteristics of unbiased Monte Carlo methods are undesirable, and the introduction of bias to the estimator can provide visually more pleasing images. The design and effectiveness of all algorithms for global illumination is reliant on their ability to handle a number of key factors, including:

- Wide range of lighting phenomenon
- High geometric complexity
- Realistic material models
- Accurate visibility queries

The trade-off of accuracy to improve performance often involves the use of approximations that reduce the robustness of the algorithm with respect to one or more of these elements. Photon mapping is a popular technique that introduces several sources of bias in an effort to improve performance and reduce per pixel variance over pure Monte Carlo methods. The novel work introduced in Chapter 5 will build upon recent advances in the photon mapping paradigm, so a thorough discussion of its theoretical background and related work is presented.

4.2 Photon Mapping

Photon mapping, introduced by Jensen [Jen96, Jen01] is a two pass bidirectional approach based on the principles of Monte Carlo light tracing, storing information carried by each light path vertex (photon) in a data structure; the photon map. Radiance measurements are obtained using a second pass, tracing a path from the camera and carrying out density estimates over the photon map to construct local flux estimates. Being based on light tracing, photon mapping is particularly effective for the generation of caustics, but with the ability to evaluate the full global illumination solution, including specular and glossy reflections, participating media [JC98] and subsurface scattering [JMLH01].

4.2.1 Photon tracing and storage

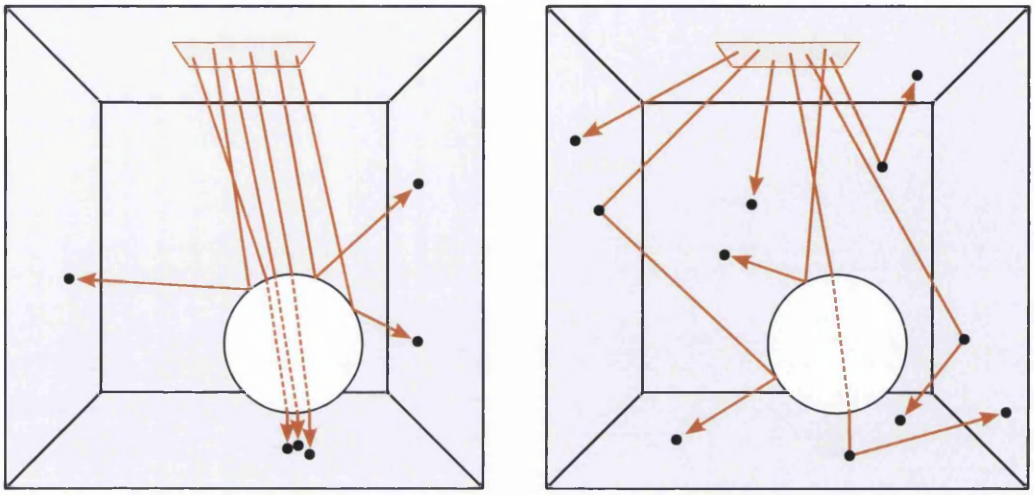
Photon tracing is performed in a similar fashion to light tracing, stochastically sampling positions and outgoing directions on light emitters according to some probability distribution and the emitter's characteristics. The now familiar path extension mechanisms and variance reduction techniques described in Section 4.1.2.

Instead of explicitly connecting each non-specular light path vertex to the camera as in MC light tracing, the vertex data is cached in a *photon map* data structure. This builds up an approximate representation of the incident illumination across the scene. Caching path vertices in the photon map allows sub-paths from the light sources to be used by multiple pixels, improving correlation and reducing per-pixel variance and high frequency noise.

Typically this first photon tracing pass can accumulate in the order of 10^4 to 10^7 photons depending on the scene complexity and desired trade-off between accuracy and performance. Caching such a large amount of data makes the representation of both photons and the photon map structure crucial. A single photon is typically represented by its position and surface orientation in 3D world space, the flux it carries and the direction from which it arrived:

```
Photon{
  float[3] position
  float[3] flux
  float[2] incident direction
  float[2] normal
}
```

The incident direction and surface normal can be represented in the global coordinate system using polar coordinates (ϕ, θ), representing the azimuth and zenith respectively, and are used to obtain and improve radiance estimates (discussed below).



(a) Caustic photon map

(b) Global photon map

Figure 4.8: *Illustrations of the photon paths that make up the caustic photon map LS^+D and global photon map $L(S|D)^*D$. All surfaces are diffuse except the sphere which is a dielectric, with refraction denoted by dashed lines.*

The flux representation is dependent on the spectral model used during rendering, generally using a single floating point value for each spectral channel, 12-bytes in the case of RGB rendering. This can also be compacted using Greg Ward's 4-byte packed format [War91] to save memory, however on modern systems it is usually not a problem and not worth the decreased accuracy and performance.

As with the particles in light tracing, it is advantageous for all photons stored in the photon map to have similar power, to reduce variance in the density estimates [Jen01]. If the flux of photons differs greatly, density estimates must incorporate a larger area of the scene during the radiance estimation pass to maintain similar levels of variance, reducing efficiency.

Applying Russian roulette at each vertex also improves the overall efficiency without additional bias, since the number of photons stored is reduced without affecting the total flux held in the photon map. This controlled increase in variance from path termination results in fewer ray computations during the photon tracing pass, alongside lower memory requirements and faster density estimates.

In many implementations of photon mapping, two data structures are utilised; a caustic photon map and a global photon map. The caustic photon map stores paths of the form LS^+D, which often produce high frequency illumination detail that is difficult to preserve using density estimation, so can require far larger numbers of photons to represent effectively. By using this additional caustic map, efficiency of the algorithm can be improved with respect to the density estimation queries. Additionally, the high density and energy of caustic photons can be problematic when mixed with sparse general illumination features, reducing the quality of the radiance estimates due to increased flux variance. Producing caustic photons exclusively can be achieved by explicitly casting photons from the light sources towards specular objects,

using projection maps [JC95], allowing the relative size of the global and caustic photons maps to be controlled more easily.

The global photon map is responsible for representing all illumination in the scene, $L(S|D)^*D$, including photons contributing to caustic illumination. Typically, this will have a more uniform distribution and lower flux density on average. As a result, the global photon map can be coarser than that of the caustic photon map, without the same loss of detail.

During the photon tracing pass, the two sets of photons are maintained as flat unordered arrays. Radiance estimation (discussed below) relies upon the representation of the photon maps for its efficiency, and often involves many millions of queries. The kD-tree is a good choice for nearest neighbour queries and provides many desirable properties including:

- Fast query performance,
- Compact storage,
- Effective handling of non-uniform distributions.

kD-trees have been discussed in Chapter 2 in the context of accelerating geometry intersections for ray tracing, but for photon mapping efficient kD-tree construction is somewhat simpler. Each photon represents only a single point in space, avoiding the need for expensive clipping or duplicate references to the data in leaves of the tree. Construction is also faster in practice, using a median-based split heuristic, resulting in each child node having roughly equivalent numbers of photons. This produces a balanced binary tree with good performance for neighbourhood searching, whose construction represents only a small part of the computation performed during photon mapping. Wald [WGS04] has investigated alternative approaches to kD-trees for photon mapping, based on an adaptation of the surface area heuristic used for ray tracing.

4.2.2 Radiance estimates

The photon map provides a discrete representation of the incoming flux Φ_i at a number of points (the photon locations) throughout the scene. For an arbitrary location in the scene an approximation of the incoming flux can be obtained by querying the photon map, and constructing the scattered radiance using the local photon data. Recall the equation for radiance scattered towards ω as an integral over the hemisphere at location x:

$$L_r(x,\omega) = \int_{\Omega_x} L_i(x,\omega') f_r(x,\omega,\omega')(\mathbf{n}_x \cdot \omega')\mathrm{d}\omega' \tag{4.6}$$

The incoming radiance L_i can be obtained from the photon map, that holds the incoming flux Φ_i of photons around x. From the definition of flux in radiometry (Section 1.5) this can be rewritten in terms of incident radiance:

$$L_i(x,\omega) = \frac{\mathrm{d}\Phi_i(x,\omega')}{\mathrm{d}\omega \mathrm{d}A \cos\theta}$$

Thus, substituting the flux estimate into the radiance integral in Equation 4.6 produces:

$$L_r(x,\omega) = \int_{\Omega_x} \frac{\mathrm{d}\Phi_i(x,\omega')}{\mathrm{d}\omega \mathrm{d}A \cos\theta} f_r(x,\omega,\omega')(\mathbf{n}_x \cdot \omega')\mathrm{d}\omega'$$

An approximation of this incoming flux at x can be obtained by integrating over the discrete representation held in the photon map. By sampling from n of the nearest photons surrounding x, the incident radiance can be approximated by a finite sum of the incoming flux $\Phi_p(x_p, \omega_p)$ for each photon p:

$$\begin{aligned} L_r(x,\omega) &\approx \sum_{i=1}^{n} f_r(x,\omega'_{p_i},\omega)\frac{\Phi_p(x_{p_i},\omega'_{p_i})}{\Delta A} \\ &\approx \frac{1}{\Delta A}\sum_{i=1}^{n} f_r(x',\omega'_{p_i},\omega)\Phi_p(x_{p_i},\omega'_{p_i}) \end{aligned}$$

where the area of the neighbourhood A is dictated by the volume occupied by the n photons around x. The incident direction ω for the BRDF and cosine factor is obtained from each individual photon in n.

Naïvely, the nearest photons are found by expanding a sphere around x. Since the scene space over which the kd-tree is built lies in 3-dimensional Euclidean space, the volume can be approximated as $A = \frac{4}{3}\pi r^3$, where r is the distance to the furthest of the n photons. In general, increasing the number of photons in the photon map provides ever increasing accuracy in the radiance computation. Higher overall density in the photon map ensures A becomes smaller due to the fixed n.

4.2.3 Rendering

After construction of the photon map, the second pass uses the radiance estimation above to evaluate the measurement function for each pixel in the image. The render pass closely follows the approach taken to MC ray tracing (as opposed to path tracing), extending camera paths into the scene until reaching a non-specular vertex, x_d. At this point, the radiance leaving x_d towards the camera is computed. This can be divided into the emitted and scattered radiance:

$$L(x,\omega) = L_e(x,\omega)\, L_r(x,\omega)$$

As with unbiased MC methods, the limitations of light paths in the presence of specular surfaces require some path space to be evaluated solely using camera based ray tracing. Namely, light sources visible directly or through specular objects; paths of the form ES^*L. Similarly, the emitted radiance $L_e(x,\omega)$ is evaluated directly without use of the photon map. The scattered radiance $L_r(x',\omega)$ can be obtained from the cached photon map data:

$$L_r(x,\omega) = L_{r,c}(x,\omega) + L_{r,d}(x,\omega)$$

where $L_{r,c}$ is caustic lighting obtained using the caustic photon map and $L_{r,d}$ is obtained from flux held by photons in the global photon map. Given that the global photon map also contains caustic photons, care must be taken to account for caustic illumination from only one of the photon maps for a given path. Thus the caustic photon map is only used when the diffuse surface x_d is visible directly or through specular interactions. After scattering through low frequency BRDFs, the illumination is soft enough that the sparse representation of caustics in the global photon map provides a more efficient alternative to querying both maps.

The caustic photon map is visualised directly, using density estimation at x_d to approximate the radiance of caustic paths using Equation 4.7. Caustics are generally of interest when photons are focussed by specular interactions, producing high photon densities thus, the radius of the search kernel can be relatively small in caustic regions, minimising the introduction of systematic bias.

The remaining $L_{r,d}$ term can be obtained from the flux contained within the global photon map. Visualising the global photon map directly can result in the gather area for such estimates to become large relative to the footprint of the pixel, due to the coarseness of the global photon map and the need to locate the n photons necessary to reduce variance. In turn this increases the risk of visual artefacts and bias in the image. As these such surfaces are directly visible, it is desirable to minimise error especially in the presence of high frequency detail; not something well handled using density estimation alone.

Instead, more accurate estimates can be acquired at additional cost, by using Monte Carlo ray tracing to extend the camera path from x_d to add an unbiased level of indirection to the radiance obtained via density estimation. This reduces the impact of noise produced due to low photon densities. Furthermore, this requires direct lighting to be computed explicitly using shadow rays as in unbiased MC methods, adding further computation but also preserving the high frequency shadow edges which the human visual system is sensitive to.

Achieving the same quality via density estimation alone would require considerably more photons, especially in the case of higher frequency BRDFs, where extensive coverage of the hemisphere over the surface would be required in addition to the surface area itself.

4.2.4 Bias, variance and consistency

Although the use of density estimation can save a considerable number of rays and improve performance, it has shortcomings. Density estimation introduces a number of sources of bias, whose elimination has been the focus of many years of research. This section will outline the cause of these problems and some key literature that aims to improve the accuracy and robustness of photon mapping.

Storing a finite number of photons provides a discrete approximation to the otherwise continuous function of scattered flux throughout the scene. The performance benefits of photon mapping are a result of the introduction of bias in place of variance, largely dictated by the size of the kernel r and its ability to reuse neighbouring sub-paths. Ideally, this trade-off would introduce only enough bias to remove the visible effects of variance. Intuitively, as the number of photons n in the gather kernel decreases, the variance of each photon p becomes proportionally more prevalent in the estimate. In terms of bias, as n increases the kernel bandwidth r also increases, thus additional bias may be introduced due to the increased distance between the photons and the gathering point. There are three main types of bias present in photon mapping that result:

- **Boundary bias** occurs in the presence of discontinuities in geometry, where the estimation kernel overlaps the boundary between these discontinuities resulting in the darkening of edges and corners, and blurring of otherwise sharp features.

- **Topological bias** is due to photons that lie on surfaces neighbouring the surface from which the camera vertex x lies. Inclusion of such photons can easily impact the radiance estimate, resulting in under or over estimation. Assuming the surface is flat across the kernel also incurs topological bias. For a curved surfaces, the area of the kernel is under estimated resulting in an increase in the estimated flux density.

- **Proximity bias** is the result of density estimation blurring the local density around the gather point, smoothing the illumination across the kernel bandwidth which can remove high frequency details such as sharp caustic detail or shadow boundaries.

A significant body of research has been introduced to tackle the problems of the variance/bias trade-off, which results from producing a radiance estimate on a 2-dimensional surface, from a 3-dimensional representation of the incident illumination. Some significant works will be summarised below.

Kernel design To reduce the types of bias mentioned above, the gather kernel can be designed to apply non-uniform weights across the photons dependent on their similarity with the gather point. The *shape* of the kernel can be modified to better suit the underlying geometry, thus providing a more accurate approximation of the volume containing the photons. The most basic spherical kernel, where $A = \frac{4}{3}\pi r^2$ in Equation 4.7 is highly inaccurate for approximating density over surfaces, since photons lie within a 2D manifold in 3D space. Topological bias can be reduced by considering the kernel in higher dimensional (non-Euclidean) space, including parameters of the local geometry around the gather point such as:

- 2D/3D Euclidean distance,
- Surface normal at x',
- Local photon orientation (ω'_p),
- Local photon distribution,

which help to modify the naïve spherical kernel to better fit the 2D manifold of the surface, improving density estimation. Applying kernel filters, such as the cone filter [Jen01], increases the influence of nearby photons, since they are likely to better represent the incident illumination at the gather point. Additionally, small changes in the kernel bandwidth can produce large changes in the radiance estimate, visible as splotches in the final image, which can be somewhat alleviated using kernel filters.

Modifying the kernel shape via simple flattening of the spherical kernel along the local surface normal to create an ellipsoid or disk shaped kernel has been shown to provide a good approximation [Jen01]. As the gather kernel becomes smaller, it can be assumed that the local surface will become locally flat, thus allowing $A = \pi r^2$ in the radiance estimate.

Further improvements to the gather kernel can be achieved by accounting for the surface normal and incident direction of gathered photons. This can reduce the influence of back facing photons that are included in the domain of the kernel due to thin objects or surfaces with high curvature.

Anisotropic kernels can also be useful making use of the local photon distribution [SSO08] or photon differentials [SFES07] to shape the kernel adaptively, reducing bias at illumination and geometric boundaries. Similarly, geometry bias can be reduced by analysis of the underlying geometry to improve the estimation of A for non-planar surface regions [HP02, TM06, HS07]

This discrepancy between the surface area measure of the kernel A and the true area encompassing the n gathered photons in the 2D surface manifold is what gives rise to additional bias in the estimator. However, a trade-off must be made between maintaining fast density estimation, and providing accurate and locally adaptive gather kernels which come at the cost of added complexity.

Bandwidth selection Choosing the size of the gather kernel can also have a significant effect on the resulting radiance estimates. Too few photons and the resulting image will exhibit high levels of noise, whilst too large a bandwidth can incur bias in the form of low frequency blurring in addition to high query costs. Bandwidth selection is an outstanding problem in many filtering contexts for computer graphics, not just photon mapping. Although effective iterative solutions are available (see [JMS96, Sch03] for an overview) these are generally too costly for application to computer graphics and/or require careful parameter tuning.

The problem of optimal bandwidth selection is an important one due to its strong influence on the resulting images. The use of adaptive approaches impacts the characteristics of the output with respect to the variance and bias of the estimator, so such techniques must be chosen carefully. This is further discussed in Chapters 5 and 6 which build upon photon density estimation and filtering kernels for image based convolution respectively.

4.2.5 Limitations of classical photon mapping

Despite the notable improvements discussed above, practical limitations still apply to such an approach. Namely, the amount of available memory limits the ability of the photon map to accurately represent the underlying continuous illumination field, pointed out by [HOJ08] among others [HHS05, HS07]. This loss of accuracy is exacerbated by density estimation in the presence of a number of features including:

- Hard illumination (eg: caustics, hard shadows),
- Small scale geometry,
- High curvature surfaces.

In addition, increasing the photon density not only increases memory consumption for the algorithm but impacts the performance of the density estimates and kD-tree traversal. Thus these penalties and the limits imposed by current memory models ensure that photon mapping as described above cannot always produce sufficient quality results for scenes with such features. A number of techniques have aimed to reduce the memory consumption of the photon map, or conversely improve the quality of the estimates that can be obtained with a finite numbers of photons. Fan et al.[FCL05] introduced Metropolis sampling to photon tracing, depositing more photons in important regions of scene space in turn producing higher quality images with the

same number of photons. Spencer and Jones [SJ09, SJ13] showed that modifying the topology of an existing photon distribution using constrained repulsion allowed the use of smaller kernels, improving rendering performance and reducing bias. Modifying deposited photons can however introduce bias into the photon map itself, in addition to the bias resulting from density estimation, and so care must be taken to respect discontinuities in the illumination. However such features may not be sufficiently represented in the existing photon map since it is already a discrete approximation.

Havran et al.[HHS05] and Herzog et al.[HS07] reorganised the photon mapping algorithm, storing intersection points of the camera paths in a tree prior to the photon shooting pass, removing the need to store photons and splatting photons to camera path hit points as opposed to gathering them.

Whilst providing improvements in the quality and performance of photon mapping, the inherent memory limitations are still prevalent in the above mentioned techniques, since storing a finite number of photons N in the map still presents a discrete approximation to the continuous function of scattered radiance.

The ability to increase the number of traced photons arbitrarily would in turn allow an unbounded increase in the photon density, improving the underlying approximation. Increasing N whilst maintaining a fixed number of photons n in the gather kernel results in a decreasing kernel bandwidth, and hence area A. Thus, bias decreases proportionally to N on average. Theoretically, as the total number of photons N approaches infinity, the estimate of the reflected radiance at a surface point x' in the scene approaches the true reflected radiance $L_r(x', \omega)$ [Jen01]:

$$\lim_{N\to\infty} \frac{1}{\pi r^2} \sum_{p=1}^{\lfloor N^\beta \rfloor} f_r(x', \omega'_p, \omega) \Delta\Phi_p(x', \omega_p) = L_r(x', \omega) \tag{4.7}$$

where $\beta \in]0:1[$. Given that an infinite number of photons are used to represent the integral, an infinite number of photons can be used in the estimate, and the solution will converge providing N^β is infinitely smaller than N. The kernel bandwidth r will become infinitesimal and converge to zero. Over an infinitesimal disk (or sphere) the effects of bias as described above are alleviated, theoretically leading to an accurate estimate of $L_r(x', \omega)$. It is this that drives the recently proposed reformulation of photon mapping into a progressive variant.

4.2.6 Progressive Photon Mapping

Progressive photon mapping (PPM) devised by Hachisuka et al. [HOJ08] provides improvements over classical photon mapping described above, removing the need to store photons whilst gradually reducing the kernel bandwidth, leading to a reduction in visible bias. Progressive photon mapping extends the reverse photon mapping and photon splatting ideas of Havran et al.[HHS05] and Herzog et al.[HS07] discussed above, storing only camera path hit points but without the use of final gather rays, and splatting photons to nearby camera hit points without additional storage requirements. The resulting two-pass algorithm has a low memory footprint, is progressive in nature and can produce final images of higher quality than conventional photon mapping.

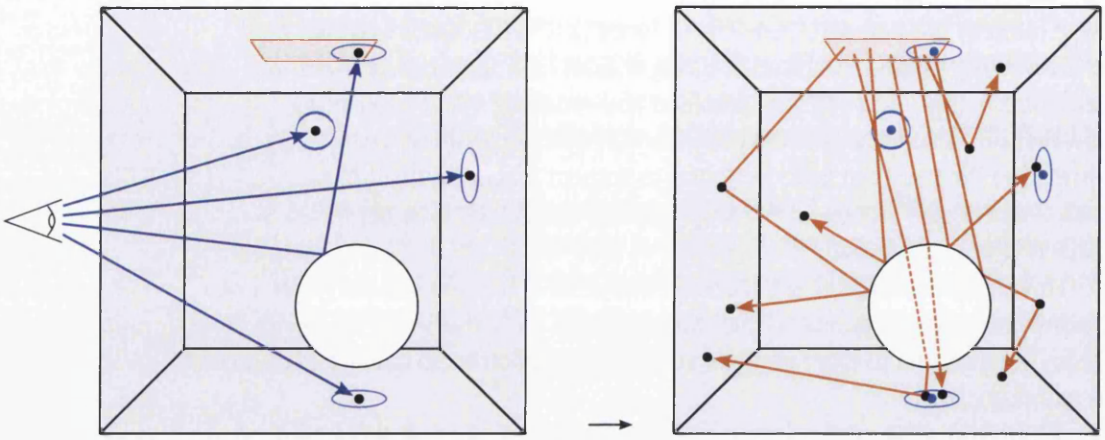


Figure 4.9: *Progressive photon mapping consists of a ray tracing pass (left), distributing hitpoints around the scene followed by a photon tracing pass (right) to compute radiance estimates.*

This progressive variant of photon mapping is an important and high impact recent development to density estimation in global illumination and forms part of the novel contributions proposed in Chapter 5. As a result it will be the focus of detailed discussion in this section to provide background and motivation for work in this thesis.

An illustration of the progressive photon mapping algorithm is outlined in Figure 4.9. An initial pass uses recursive ray tracing to identify a finite set of points on non-specular surfaces visible through each pixel. Subsequent passes then trace a set of photons, storing a photon for each non-specular surface, building up a photon map akin to the initial pass found in the original PM algorithm. After each photon tracing pass, a photon map is constructed and density estimation is performed around each of the stored camera hitpoints. After a set number of photons have been traced, the current photon map is used to refine the flux estimate and kernel radius for each hit point ready for the next pass. Once the radiance estimate at each hitpoint has been updated, the photons for the current pass are discarded and a new photon tracing pass can begin, maintaining the combination of constant memory usage and reduced bias that eluded previous approaches.

This process is analogous to averaging multiple independent photon maps, where each iteration trades bias for variance as the kernel bandwidth r is reduced over time, tending towards zero. This progressive reduction in kernel bandwidth and overall increase in photon contributions is what provides convergence to a consistent solution with reduced variance and bias. The resulting PPM estimator aims to remove the upper limit upon the number of photons contributing to the radiance estimate at each hitpoint, attempting to fulfil the convergence property expressed by Equation 4.7.

4.2.6.1 Camera hitpoints and statistics

Acting as gather points for density estimation, each camera hit point from the initial pass stores sufficient data to perform both density estimation and radiance measurements for the associated pixel:

```
Hitpoint{
  float[3]      x              Position
  float[3]      n_x            Surface normal
  float[3]      ω              Ray direction
  BRDF*         f_r            BRDF at x'
  float[2]      P(i,j)         Target pixel
  float[3]      T(x̄)           Path throughput to x'
}
```

In addition to the hitpoint location and surface normal for density estimation, the camera ray direction ω and BRDF at x are also included to compute the scattered radiance. The path throughput $T(\bar{x})$ from the camera to x is cached at the hitpoint along with the paths origin on the image plane, to allow radiance contributions to be added directly to the image buffer. Statistics are also maintained to track the current kernel radius r and accumulated photon count N, which allow for the progressive reduction of kernel bandwidth and correction of the flux estimate after each iteration (discussed in detail below):

```
Statistics{
        float   r       Current kernel radius
        int     N       Accumulated photon count
        colour  τ       Accumulated unnormalised flux
}
```

Though allowing for arbitrary numbers of photons, the initial pass restricts PPM to evaluating radiance at a finite number of visible points. This becomes problematic for example in the anti-aliasing of geometric and texture edges where the points need to be well distributed and numerous enough to resolve fine detail. Additionally, there is no support for distributed rendering effects such as glossy surfaces and depth of field.

Stochastic hitpoint generation Stochastic progressive photon mapping [HJ09] (SPPM) adds the ability to compute radiance estimates over a visible region, as opposed to a set of discrete points, whilst maintaining the desirable properties and progressive nature of the original algorithm. In general, this is achieved by adding a distributed ray tracing pass between photon tracing passes in order to obtain new camera hitpoints lying inside the visible region of each pixel. This enables flux to be estimated over a different visible location at each iteration x_i, which over successive iterations cover the entire space S visible from a given pixel, where $x_i \in S$ for all i. As a result statistics are maintained for each pixel in SPPM representing the current kernel radius and the total flux contributed from pervious iterations, as opposed to maintaining a set of statistics for each hitpoint as in PPM.

By changing both the camera hitpoints and the photon distribution at every iteration stochastically, SPPM can provide radiance estimates for complex integrals in a progressive manner, without the limitations of previous photon mapping techniques. The remainder of this section

will discuss the stochastic formulation of the algorithm in further detail, since it is subsumes the original PPM approach and is significant to the novel work proposed in Chapter 5. For further details of the differences and derivations of the stochastic variant from the original approach see [HJ09]. The notation used in the context of (S)PPM follows that of the original authors [HOJ08, HJ09].

4.2.6.2 Radius reduction

The key concept behind (S)PPM is the reduction of the kernel bandwidth to decrease bias. Instead of performing a nearest neighbour search for fixed n, the density estimate is produced from photons that lie within a fixed radius r of the gather point. For a given location x the photon density of photons within a fixed radius r can be calculated simply as:

$$d(x) = \frac{n}{\pi r^2}$$

assuming a uniform kernel filter and a planar surface at x. For the initial iteration i of the algorithm, the photon density can be computed as in the classic photon mapping approach using the number of photons located in the current gather radius. In discarding the photon map, the density at each subsequent iteration must consider the total number of accumulated photons $N_i(S)$ collected from all previous iterations up to i over the set of surface points S visible through the pixel. Thus the density must be updated based on the number of new photons $M_{i+1}(x_i)$ collected at the hitpoint $x_i \in S$ for the new iteration $i+1$:

$$d_{i+1}(S) = \frac{N_i(S) + M_{i+1}(x_i)}{\pi R(S)^2}$$

However to reduce bias, the (S)PPM algorithm must also reduce the kernel radius. Given an ideal photon distribution, for a region of uniform illumination any density estimate performed should ideally compute the same photon density for that region, regardless of the kernel bandwidth. To maintain a consistent estimator (as in Equation 4.7), the number of photons that contribute to the radiance estimate for each pixel region must increase towards infinity, otherwise flux contributions would be lost and would have no effect on the final estimate [3]. Thus maintaining constant density requires the gather area to decrease proportionally to the number of photons that are accumulated. A global parameter α is used to dictate the rate of radius change at each iteration based on $M_i(x_i)$:

$$N_{i+1}(S) = N_i(S) + \alpha M_i(x_i) \tag{4.8}$$

Since convergence only occurs due to the decreasing kernel radius, the parameter α provides a balance in the trade off between bias and variance. Given that S has a kernel radius of $R_i(S)$ at iteration i the new radius $R_{i+1}(S)$ must ensure that the new area $\pi R_{i+1}(S)^2$ decreases proportionally to the change in $N_i(S)$ [HOJ08]:

$$R_{i+1}(S) = R_i(S)\sqrt{\frac{N_i(S) + \alpha M_i(x_i)}{N_i(S) + M_i(x_i)}} \tag{4.9}$$

[3] Intuitively, if the number of photons did not increase this would be equivalent to averaging the estimate at each iteration, which does not reduce bias effectively

Along with the current kernel radius and number of accumulated photons, each pixel (in the case of SPPM) tracks the currently accumulated flux over S, which must similarly be adjusted.

4.2.6.3 Flux correction and radiance computation

Contrary to classical photon mapping, each pixel's flux is stored *unnormalised* with respect to the gather area and the number of photons traced, since both vary during rendering. For each set of new photons collected over the gather kernel with area $\pi R_{i+1}(S)^2$ and number of photons $M_i(x_i)$, their combined unnormalised flux can be calculated:

$$\Phi_i(\mathbf{x}_i, \omega) = \sum_{p=1}^{M_i(x_i)} f_r(x, \omega, \omega_p)\Phi_p(\mathbf{x}_p, \omega_p)$$

Notice that the flux of each photon is multiplied by the BRDF at x. This new flux needs to be combined with the flux accumulated over all previous iterations, in order to satisfy consistency. Progressively reducing the kernel radius means that photons that lie within R_i may no longer lie within the kernel defined by R_{i+1}, requiring an adjustment in the flux estimate from the previous iteration to account for photons that now lie outside the new radius. The assumption of constant photon density is maintained and the refined flux is scaled by the change in area of the kernel:

$$\tau_{i+1}(S, \omega) = (\tau_i(S, \omega) + \Phi_i(x_i, \omega)) \left(\frac{R_{i+1}(S)}{R_i(S)} \right)^2 \tag{4.10}$$

where $\tau_i(S, \omega)$ is the unnormalised flux estimate for iteration i scattered towards the direction of the camera ω from visible surfaces in S. If the area $\pi R_i(S)^2$ of the gather kernel is reduced over a photon distribution of constant density, the number of photons will decrease proportionally. In turn assuming each photon carries the same flux on average, the total unnormalised flux contributing to the radiance estimate should be decreased by the same proportion, since the current $R_i(S)$ is used for flux normalisation.

A radiance estimate can be obtained after any iteration of the algorithm by normalising $\tau_i(S, \omega)$ using the shared statistics. Following the normalisation process of conventional photon mapping, the radiance scattered towards the camera at ω can be calculated for the shared pixel flux as:

$$\begin{aligned} L(S, \omega) &\approx \frac{1}{A} \sum_{p=1}^{n} f_r(x, \omega'_p, \omega)\Delta\Phi_p(x_p, \omega'_p) \\ &\approx \frac{1}{\pi R_i(S)^2} \frac{\tau_i(S, \omega)}{N_e(i)} \\ &= \lim_{i \to \infty} \frac{\tau_i(S, \omega)}{N_e(i)\pi R_i(S)^2} \end{aligned} \tag{4.11}$$

where $N_e(i)$ is the total number of photons emitted up to and including iteration i. In the limit a radiance measurement is obtained that is in theory equivalent to solving the radiance integral of the surface area visible from that pixel using Monte Carlo integration. In practice this is not the case and is discussed further in Chapter 5.

Convergence and extensions The progressive reformulations of photon mapping are initially plagued by the same geometric, topological and proximity bias as classical photon mapping, dependent heavily on the radius of the gathering kernel [Jen01].

Hachisuka et al.[HJJ10] discuss error estimation in PPM, and provide a means to estimate the relative levels of variance and bias forming the overall estimator error, and show how it can be used as a stopping criteria. Knaus and Zwicker [KZ11] similarly present a rigorous analysis of the PPM algorithm's properties, establishing that its variance decreases with $O(1/N^{\alpha})$ and bias with $O(1/N^{1-\alpha})$. Thus the choice of α is important to determine the behaviour of PPM and the convergence of overall error. The formulation of PPM described by Knaus and Zwicker further shows that the convergence of the algorithm is independent of the number of photons gathered and the local photon density. Thus the collection of such statistics as originally presented by Hachisuka is unnecessary. In turn this permits each iteration of their probabilistic approach to be run in parallel as the combination of individual photon maps, since the radius of each iteration and the resulting radiance estimate is independent of all previous iterations.

The error convergence of $O(1/N^{\alpha})$ with $0 < \alpha < 1$ results in poorer convergence characteristics than Monte Carlo path tracing approaches in which error is reduced to the order $O(1/\sqrt{N})$. The density estimation and photon caching properties of (S)PPM make for increased sampling performance, and allow it to handle a number of scenarios that cause difficulties for unbiased path tracing approaches. This forms part of the motivation behind the novel work proposed in the following chapter, and a comparison between the two approaches and their limitations will be discussed in detail.

Hachisuka and Jensen [HJ11] present an adaptive photon tracing scheme that allows the distribution of photons to be adjusted using an importance function based on photon visibility. This allows a higher photon density in visible regions, improving efficiency for close ups of large or highly occluded scenes where a large percentage of the deposited photons do not contribute to the image. However, their approach uses a binary visibility function that does not take into account the importance of photons with respect to their contributions. Thus, a large number of photons may still be deposited in regions distant from the camera that contribute minimally to the image.

4.3 Alternative Monte Carlo methods

Alternative approaches to light transport have been devised that also rely on the introduction of bias to Monte Carlo methods with the aim of reducing variance. Many of these techniques are less robust than the path tracing and photon mapping techniques described above and are often limited in the range of lighting phenomenon they can solve, or the materials that they are able to handle. This makes design difficult for artists and designers that rely on accurate material rendering or complex geometry, and can require additional effort to produce the desired effects, requiring familiarity with the algorithms or their results in order to counteract the bias and inconsistencies introduced by such techniques.

Many of these such as instant radiosity [Kel97] and numerous derived techniques provide rapid convergence for diffuse and low frequency lighting but cannot effectively handle

specular materials and caustics. Virtual point light approaches work in similar fashion to photon mapping, consisting of a photon tracing pass, caching illumination around the scene using virtual point light sources (VPLs). The second rendering pass replaces density estimation with point to point gathering using shadow rays or maps to each VPL. This is a popular approach for interactive rendering due to its efficient path reuse, but can introduce significant image artefacts due to the normalisation parameters used to compute contributions from the VPLs (see for example [KK04]). Many extensions to the approach improve robustness and enable the handling of non-Lambertian surfaces [HKWB09] or include volumetric effects [NNDJ12b, NNDJ12a, WKB12]. To obtain good results without artefacts, a large number of VPLs are required. To maintain performance and scalability with numerous VPLs research has focussed on clustering [DGR+09] or hierarchical approaches [WFA+05] to provided bounded approximations. A comprehensive reference to point light based methods and related techniques, in addition to those suitable for interactive global illumination can be found in the recent survey of Ritschel et al.[RDGK12] and SIGGRAPH course by Křivánek et al.[KHA+12], but detailed discussion is outside the scope of this thesis. Point light based techniques are becoming increasingly competitive with path tracing and photon mapping approaches for more complex scenes as the types of phenomenon they can handle is expanded (see for example [WKB12]). Nevertheless, they are still unable to handle importance features such as caustic lighting and can incur noticeable bias on complex materials in order to remain efficient, making them inappropriate where accuracy and robustness are required.

4.4 Extending unbiased Monte Carlo methods

Unbiased path tracing methods provide an effective solution to high quality global illumination. Having been at the forefront of state of the art global illumination research for many years, a significant number of techniques have resulted, focussed on addressing its limitations and improving efficiency.

The work proposed in this thesis follows the philosophy that extending robust unbiased techniques is an effective and highly future orientated approach, as discussed in the introduction of this thesis. Biased estimators, or those which lack fundamental robustness with respect to scene and illumination features do not necessarily converge to the correct result and as such, they can require considerable research attention to achieve the desired capabilities. Continuation of the current trends in hardware clock speeds and parallelism will likely give rise to faster convergence for existing biased approaches, but still produce results of insufficient quality [4].

On the contrary, unbiased techniques can already provide high quality results and future hardware developments will thus naturally result in improved efficiency without substantial algorithmic change.

Adding and extending unbiased techniques via the addition of approximative algorithms can address intermediate noise caused by variance, while producing high quality results in the limit. It is desirable that these intermediate results are comparable to those of more sample-

[4] Inevitably, if current trends continue high quality unbiased rendering will become real-time, potentially eliminating the need for biased methods altogether.

efficient biased techniques. This also preserves the desirable practical properties of the underlying unbiased approach, and the potential of such methods to adapt easily to new hardware.

The final background section of this thesis will discuss a number of paradigms that build upon Monte Carlo methods, focussing on improvements to unbiased path tracing. The aims of such approaches are to improve convergence, visual quality of intermediate results or to produce ultimately biased results using the path samples from unbiased integrators. Recent techniques in each category are discussed along with their strengths and weaknesses to highlight open problems.

4.4.1 Cache-based importance sampling

As discussed in Chapter 3, importance sampling techniques attempt to maximise a sample's contribution to the estimator with respect to its importance to the originating sensor or emitter. However such a greedy importance sampling approach can result in a poor approximation of the illumination since it ignores the incident radiance or sensor response during the construction of camera or light paths, respectively. Obtaining a coarse approximation of the integrand that accounts for some of this missing importance can improve the accuracy of the probability distribution functions used for importance sampling. A number of techniques employing this approach to reduce sample variance are discussed in this section.

Path mutation Metropolis Light Transport [VG97] and Energy redistribution path tracing [CTE05] discussed in Chapter 4 build up an importance function over time via the local exploration of path space. This improves convergence in scenes with difficult lighting and occlusion since connections established between the camera and emitters can be modified, leveraging the generally high coherence of the illumination integral.

In contrast to explicit cache based approaches, path mutation schemes cannot share importance information globally, and upon leaving a region of path space (to ensure ergodicity) such importance information is lost. Caching even a sparse representation of local importance can help guide the rendering process at arbitrary path vertices irrespective of the current state of the renderer.

Particles as importance indicators Instead of relying on BRDF importance sampling alone to generate new samples, information obtained during the rendering process can be used to determine areas of the integrand that hold high importance, guiding future sample generation. Several methods have used a pre-pass based on particle tracing to generate an approximation of the incident illumination that can be queried during the MC rendering pass.

Jensen [Jen95] and Hey [HP01] utilised the photon map to provide illumination information to aid importance sampling in path tracing. An initial pass creates a sparse photon map, creating a coarse representation of the incident radiance. During path tracing, the photon map is queried at each path vertex to construct a discrete PDF over the unit hemisphere. This distribution function is used to provide radiance based importance for sampling high contribution path space.

Budge et al.[BAJ08] adopt a similar two pass approach, specifically for rendering caustics. After creation of the photon map, a clustering procedure instead groups nearby photons and

each cluster is ranked according to its flux density. During rendering, an existing unbiased VPL method [KK04] is used to compute non-caustic illumination and the photon clusters used to improve path tracing.

Georgiev et al.[GKPS12] build an unbiased sampling technique in the context of VPL rendering, constructing a set of four increasingly conservative PDFs, as a form of defensive importance sampling (recall Section 3.3), to avoid excessive variance around discontinuities. These PDFs are then shared between neighbouring samples to improve correlation.

Pre-passes such as photon tracing introduces additional computation, as they do not contribute directly to acquired radiance, and strip away the progressive nature of MC path tracing. Additionally, the number of photons necessary to obtain a good representation of the illumination is not known, easily resulting in insufficient data to employ importance sampling, or inefficient oversampling.

Adaptive importance caching Alternatively, the radiance contributions of previous samples can also be utilised during rendering to guide subsequent samples, without the need for pre-processing. Furthermore, all samples contribute directly to the estimator as radiance measurements.

Cline et al.[CAE08] introduce a strategy that constructs and samples a discrete PDF to guide random variables used for path construction, based on the integral of neighbouring pixels. However their approach is only effective up to the first non-specular vertex, unsuitable for progressive rendering, and can introduce high contrast artefacts where pixel similarity breaks down, such as geometric edges, since it relies on a local image based model.

Bashford-Rogers et al.[BRDC12] instead build up a view independent object-space cache of important directions at discrete points in the scene. Using compact cosine lobes allows easy updates during rendering, avoiding the requirements for a pre-process, and ensures all samples directly contribute to the final image.

Although providing guidance for path generation, care must be taken with these techniques to avoid excessive variance where changes in the integral are not sufficiently represented by the discrete PDFs. In addition, such methods do not solve the underlying difficulty of initially sampling high frequency or caustic illumination. Implicit light source connections obtained via specular paths are highly sensitive to even small path changes, and can lead to increased variance even in the presence of good integrand estimates. The characteristic spiked noise observed in MC integration can be reduced but not eliminated, as the underlying cause of the problem still remains; sampling high energy paths that have low probability densities.

4.4.2 Adaptive rendering

Adapting sample placement can help to reduce variance and error locally, providing improved convergence for a particular integrator. In order to provide visually pleasing images, it is desirable to achieve consistent convergence across all pixels of the image, such that the visible error in each pixel is the same. Distributing samples across pixel integrators according to their perceptual or statistical error can provide more uniform convergence, attending to undersampling whilst avoiding oversampling pixels with low error.

Image based approaches Early work on image based adaptive sampling were established by Mitchell and Netravali [MN88] to tackle aliasing, and Bolin and Meyer [BM98] who used a perceptual model to produce a distribution map over the image to guide pixel sampling. Recently, a number of papers have focussed on interleaving adaptive sampling and filtering [ODR09, RKZ11, RKZ12, LWC12] reducing residual error and a promising avenue for future progressive rendering methods since they filter existing pixel radiance to reduce visible error, and target sampling to reduce the residual error left in the filtered image.

Image based adaptive sampling can be effective on its own for reducing perceptual error in low complexity scenes. As the pixel integrands becomes more difficult to sample stochastically, such techniques lose efficiency due to their brute force approach to variance reduction by tracing additional samples. Combining these approaches with cache-based importance sampling and/or object-space adaptive techniques such as Markov-chain Monte Carlo (as explored by Hoberock and Hart [HH10]) is promising and could be an interesting area of study for further research.

Multi-dimensional approaches Hachisuka et al.[HJW+08] introduced a multi-dimensional adaptive sampling technique, storing samples in a high dimensional kD-tree and using Euclidean distance to estimate error in this extended path space. They can cater for arbitrary lens effects during ray tracing, but their sample storage does not scale well to suit the complexities of global illumination and sample reconstruction can become costly.

Drawbacks Adaptive rendering techniques can be advantageous for indicating under sampled regions of an integrand, and be further enhanced by perceptual metrics if appropriate (see for example [RPG99]). The main drawback of adaptive techniques is that regions can only be recognised as under sampled once the source of the variance is found. Relying on pixel variance or error metrics can result in under sampling regions with difficult lighting, if the path space that contributes to the pixel has not been properly explored. In the worst case this can produce significant bias if used without ensuring adequate coverage of the integrand, as investigated by Kirk and Arvo [KA91].

4.4.3 Sample reconstruction

Recent methods such as the work by Sen and Darabi [SD12] and Lehtinen et al.[LAC+11] rely on identifying statistical relationships between Monte Carlo samples to remove noise. Typically a small number (8-16) of Monte Carlo samples are rendered storing the spatial, texture and radiance data associated with each sample. Statistical analysis of this data is performed to determine the dependence of each samples radiance on the random parameters used to construct the Monte Carlo samples. These dependencies are then used to determine the individual samples that are effected by noise and produce filter weights as input to a cross-bilateral filter to minimise their impact on the image. Whilst providing impressive results and handling large dimensional problems, they do not scale well with increased input samples. To capture more difficult lighting phenomena such as glossy reflections and caustics higher sample counts are often needed, making such reconstruction techniques inappropriate in such situations.

Li et al.[LWC12] recently presented an adaptive sampling and filtering approach building on the work of Sen and Darabi. They reduce memory consumption by generalising towards a per pixel scale, storing the mean and variance for samples of the pixels for each dimension (as opposed to the data of each sample). Additionally, they provide adaptive sample placement over the image guided using an existing estimator from statistics to estimate filter error. Though reducing memory usage and thus allowing more samples to be rendered prior to reconstruction, no quantitative data is presented.

Although providing impressive image quality from low sample counts, the sample processing stage for these reconstruction techniques can take in the order of minutes to compute, potentially making it infeasible for the rendering of fast previews or implementation into a progressive rendering system.

The novel technique proposed in Chapter 6 of this thesis is motivated by the use of vertex data to improve filtering, but to do so in a memory efficient way for an arbitrary number of samples, while suitable for inclusion into a progressive rendering framework.

Irradiance caching The high dimensionality of light transport is suggestive of cache based solutions that pre-compute some aspect of the domain and reuse it to improve efficiency.

Ward et al.[WRC88, WH08] introduced irradiance caching, storing a sparse set of samples representing the irradiance at surface locations. An image is rendered by extrapolating between cached sample to reconstruct the irradiance at the surface point being shaded. Artefacts are reduced by using the derivatives of position and surface orientation to better predict irradiance changes during extrapolation. However, problems still arise due to the density of the irradiance cache, especially in regions of non-planar geometry. Like any discrete approximation, discussed earlier in the context of photon mapping, storage costs for cache based solutions restrict their accuracy to properly reconstruct the illumination from a set of discrete points.

Kontkanen et al.[KRK04] use image-based irradiance filtering to reduce noise, based on the assumption of filtering low frequency irradiance. These assumptions restrict the techniques usefulnesss since high frequency detail is common when dealing with non-Lambertian materials. Self shadowing and illumination gradients caused by geometry are not easily preserved, and rely on additional costly 'feeler' rays to detect and preserve local geometric changes.

Radiance caching and filtering algorithms have also been proposed [GKB09, KGPK05] based on the ideas of irradiance caching, with the inclusion of spatial and direction components to guide filtering. However all these algorithms either rely on low frequency illumination or BRDFs, or are not applicable to general light transport.

4.4.4 Noise removal

Monte Carlo estimators are commonly associated with their distinctive white speckling. Using importance sampling based on BRDF and cosine sampling, although generally reducing variance can exacerbate the presence of high intensity speckled noise where the expected sample distribution does not accurately present the actual measured radiance field.

Once introduced into an image, such noise can be highly persistent and difficult to remove. Applying adaptive rendering techniques based on pixel variance targets sampling towards pixels exhibiting such noise, but this can be a highly inefficient technique. Due to the rarity of

such samples in the importance sampling distribution, thousands of samples can be required to offset their weighted contribution.

A more effective approach can be to remove (or *reject*) these low density samples temporarily until they become 'normalised' with respect to the distribution, when they can be reintroduced without increasing variance.

DeCoro et al.[DWR10] employ a sample rejection technique based on the local sample density in a 5D joint image-colour space, delaying the addition of samples that do not conform to the current distribution of the neighbourhood. This allows objectionable noise to be removed from scenes with non-specular materials and high frequency lighting, without the blurring and artefacts commonly associated with filtering. Rejected samples can then be re-evaluated and added to the solution when the local density becomes sufficient, remaining unbiased in the limit. Using density estimation performs admirably in well converged areas, where there is a good sense of the normal distribution of sample energies. For sparsely sampled regions it can however have detrimental effects due its harsh rejection properties, resulting in significant energy loss and removal of illumination clues. The relative luminance of samples in the same pixel are not taken into account, removing isolated samples that may not be responsible for noise due to the presence of those with higher energy.

Pajot et al.[PBP11] also build on density estimation to tackle bright spot removal, based on per-pixel 1D luminance distributions of the observed samples. This improve performance over the method by DeCoro et al., reducing the 5D k-Nearest neighbour density estimates to a 1D per-pixel kernel estimates which incur lower overheads. Conversely to DeCoro et al., neighbourhood similarities are not accounted for, leading to the rejection of samples that help represent important illumination features such as caustics, which are seen as outliers with respect to the distribution of a single pixel.

Both approaches operate over the sample radiance; the aggregated radiance of the contributions from each vertex of a path. This results in the rejection of additional samples, since high intensity noise from a single vertex will typically result in the rejection of the correlated set of radiance. Higher rejection rates result in additional bias, and darkening of the image which is especially noticeable at low sample counts.

Part of the novel work in Chapter 6 aims to address some of these issues, providing an efficient approach to remove high intensity noise with low rejection rates to minimise bias. Additionally, the new approach aims to provide a pre-processing step to convolution based image filtering.

4.4.5 Image based filtering

A vast swathe of techniques for image de-noising have been developed (see for example [BCM05a] for a review), commonly revolving around computing the output value of a pixel using the weighted sum of pixels within a neighbourhood in effort to preserve the underlying signal. Here a number of notable and relevant techniques are discussed to provide an overview of the current literature.

Suykens and Willems [SW00] use per-sample adaptive kernel widths to distribute radiance across pixels during progressive rendering. Kernel widths are dictated by the path densities of each sample, transforming noise to lower frequencies which are less objectionable. This

comes at the cost of blurring image features as it does not consider geometric or texture edges, distributing samples over strong discontinuities and losing detail.

An interesting interactive filtering approach has been developed by Bauszat et al.[BEM11] which pertains to be more robust to outlying samples and geometric edges. Instead of filtering the illumination directly, they seek to find a transformation mapping the underlying local geometry of a pixel to the noisy illumination data. This is then used to preserve the shape of the function over the geometry (dictating edges and gradients), but following the shape of the illumination input. Results provide effective noise reduction with high performance and effective geometry anti-aliasing. This approach is currently limited to diffuse, near-Lambertian BRDFs and omits direct lighting making it unsuitable for general global illumination. Only edges in the geometry buffer are used to fit the transformation function, hence features of the indirect illumination are not preserved.

Bilateral and Cross-Bilateral filtering Many modern image based filtering techniques used in rendering stem from methods such as the popular bilateral filter [TM98] and subsequent cross-bilateral filter [ED04, PSA+04]. The principle of the bilateral filter is to extend the idea of range Gaussian convolution across the 2D image space, but to incorporate additional dimensions derived from the value difference between pixels to restrict blurring across edges.

A pixel p is denoised by computing a weighted average of the pixels q in a neighbourhood. In Gaussian convolution this neighbourhood is defined over the pixel coordinates in 2D image space alone, producing the convolution operator:

$$GC[I]_p = \sum_{q \in S} G_\sigma(\|p-q\|)I_q$$

where $G_\sigma(\|p-q\|)$ denotes the value for q from the 2D Gaussian kernel centered around p, as described by:

$$G_\sigma(x) = \frac{1}{2\pi\sigma^2} e^{-\frac{\|p-q\|}{2\sigma^2}}$$

where σ is the kernel bandwidth (in this case the image space neighbourhood) and e is Euler's number. Therefore as pixels become more distant in the neighbourhood, their influence with respect to the filtered value of p is reduced.

The bilateral filter extends this concept by extending the neighbourhood to include pixel intensities. As a result the influence of q on the resulting value of p is reliant on two kinds of weights; the 2D Gaussian as above and the distance between p and q in the pixel intensity domain. Thus, the resulting bilateral filter value BF of pixel p can be calculated:

$$BF[I]_p = \frac{1}{Z_p} \sum_{q \in S} G_\sigma(\|p-q\|)G_{\sigma_r}(I_p - I_q)I_q$$

where Z_p is a normalisation constant computed from the sum of $G_\sigma(\|p-q\|)G_{\sigma_r}(I_p - I_q)$ for all q, to ensure energy preservation across the filter kernel. As a result of this additional range term, pixels that vary greatly in intensity are distant in the new neighbourhood, and thus their influence on the p is reduced by the kernel G_{σ_r}. The notation for bilateral and convolution based filtering in this thesis is based on that used in [PKT09], due to its clarity.

The cross-bilateral filter (CBF) also known as the joint-bilateral filter builds upon the standard bilateral filter, but instead utilises data from additional images or data sources to provide further dimensions to restrict the convolution filter, as opposed to relying on pixel intensities of the input image that may be corrupted by noise. Thus the computation is much the same, but the filtered output of I_p is reliant on the distance between p and q in the neighbourhood of the range buffer RB_p:

$$CBF[I,RB]_p = \frac{1}{Z_p}\sum_{q\in S} G_\sigma(\|p-q\|)G_{\sigma_r}(RB_p - RB_q)I_q$$

Eisemann and Durand [ED04] and Petschnigg et al.[PSA+04] first applied this to the problem of noise reduction in low light photography, using a photograph taken with a fill flash as a range buffer to preserve edge detail in the image based on the pixel intensities. Thus, the pleasing tones of the soft ambient lighting are preserved, without blurring the hard edge detail.

In the context of rendering, this allows a low variance buffer to dictate the dimensions and range of the filter whilst de-noising the illumination that is the source of high variance in the estimator. Commonly, range buffers are composed of surface normal orientations, depth buffer information and visible texture detail for each pixel extending the CBF to more complex neighbourhoods that better describe the continuity of the images features. Extension of the CBF to take advantage of multiple inputs is simply a matter of computing additional distances between the intensities for each range buffer RB:

$$\begin{aligned} CBF[I,RB^0,...,RB^n]_p \;=\; & \frac{1}{Z_p}\sum_{q\in S} G_\sigma(\|p-q\|)G_{\sigma_r}(|RB_p - RB_q|) \\ & ...G_{\sigma_r^{n-1}}(|RB^{i-1}(p) - RB_q^{n-1}|)G_{\sigma_r^n}(|RB^n(p) - RB_q^i|)I_q, \end{aligned} \quad (4.12)$$

Techniques based around the (cross) bilateral filter have shown to produce good results for MC rendering (see for example [XP05, SKBF12, SD12, LWC12]). This principle of the cross-bilateral filter is built upon in the novel techniques presented in Chapter 6, and implementation, limitations and comparisons will be discussed there. More discussion on bilateral filtering and an overview of its related work and applications can be found in [PKT09].

Dammertz et al.[DSHL10] successfully modify the à-trous wavelet transform, adding filter weights derived from geometry and input sample based stopping functions into a cross-bilateral filtering framework. Despite largely respecting both geometric and illumination edges, the à-trous filter can lose fine detail especially in textures, where reliance on the sparse input does not restrict the filtering sufficiently. Their filter also presents noticeable artefacts in the form of ringing around contours of illumination gradients.

Schwenk et al.[SKBF12] use a perception based blending operator, relying on pixel variance to combine path traced input samples with a filtered image, producing an unbiased algorithm in the limit. It further allows them to use the higher quality cross-bilateral filter for fast previews which can be too slow for interactive rendering. Illumination edges and glossy reflections however become blurred due to the ignorance of illumination features, the main problem addressed in Chapter 6. Their blending approach is largely orthogonal to the filtering technique utilised for noise reduction and is in the spirit of the desire for high quality rendering taken in this thesis, making it of interest for future integration with the proposed work.

General image denoising Recent techniques [RKZ12, KS13] have made use of image denoising filters to handle general Monte Carlo noise irrespective of complex effects. The non-local means [BCM05b] filter is one such popular technique, based on the bilateral filter but operating over the similarity of pixel neighbourhoods as opposed to individual pixels. Although effective and backed by a significant amount of prior research in computer vision, additional data available from the renderer is not used that could improve the effectiveness of the filtering and preserve further detail. As a result such techniques are often limited to removing more subtle low intensity noise and for cleaning up the appearance of final renders as opposed to providing previews and noise reduction for sparsely sampled regions.

4.5 Conclusion

This chapter supplied a thorough review of the numerous paradigms developed in the quest for improving light transport synthesis. Both a theoretical and implementation orientated discussion of major state of the art techniques has been carried out including recent extensions. Monte Carlo path tracing algorithms are capable of rendering a broad range of materials and lighting phenomenon. High variance in the presence of complex effects can however reduce the visual quality of results. A wealth of prior research from the mathematical disciplines has provided a number of techniques such as importance sampling and Russian roulette that have become crucial improvements. Bidirectional path tracing and Metropolis light transport further extend the effectiveness of path tracing techniques, allowing better estimates in highly occluded scenes or those with difficult lighting conditions.

Photon mapping, a popular alternative paradigm, is capable of handling complex illumination, particularly caustic lighting and specular materials. Caching a discrete flux representation in the form of photons allows variance reduction via the reuse of samples and path correlation. Bias introduced as trade off for such benefits has been discussed, along with its causes and recent avenues of improvement including progressive photon mapping. This forms an important introduction to the novel work in the following chapter of this thesis, which combines path tracing and photon mapping in a complimentary progressive formulation, reducing overall variance and bias.

Finally, a summary was provided of extensions to Monte Carlo methods that have been previously proposed to tackle the variance and noise resulting from stochastic sampling. Cache based methods, techniques for adaptive rendering and sample reconstruction from sparse signals have all been explored in recent literature and an overview has been provided to represent the plethora of techniques available, and to allow insight into the avenues that need attention as further motivation for this thesis. Noise removal and image based filtering algorithms realise the insights and proposed techniques in Chapter 6, which removes noise during progressive rendering utilising a sample clustering scheme based on path vertex data, and preserves illumination features.

Chapter 5

Mixing Monte Carlo and Progressive Rendering

Contents

5.1 Introduction **83**
5.2 Motivation **85**
5.2.1 Limitations of unbiased Monte Carlo methods 86
5.2.2 Photon Mapping for caustic illumination 88
5.2.3 Hybrid approaches 90
5.3 Overview **92**
5.4 Path space separation and filtering **93**
5.5 Path tracing **94**
5.6 Caustic evaluation **97**
5.6.1 Photon tracing 98
5.6.2 Photon gathering 100
5.7 Results and discussion **101**
5.7.1 Cornell Box 103
5.7.2 Metallic Ring - difficult caustics 104
5.7.3 Shapes - Reflective open environment 105
5.7.4 Quantitative comparisons 108
5.8 Limitations and future work **110**
5.9 Conclusions **114**

5.1 Introduction

In the previous chapters of this thesis, a number of key approaches for solving the light transport problem, either in its entirety or in part, have been discussed. All such methods have

their individual strengths and weaknesses, and consequently it can be advantageous to combine them so as to maximise the merits of each constituent algorithm, whilst minimising the impact of their failures. When drawing upon existing algorithms for global illumination, care must be taken to ensure each element of light transport is accounted for, so that a full global illumination solution can be produced. Typically, some techniques can handle highly diffuse environments well, but struggle when presented with complex specular effects. In developing hybrid techniques, additional care must be taken that each class of light transport is accounted for exactly once, so that additional bias is not introduced that may favour particular illumination features. An effective method of achieving this is to ensure that each type of light transport, characterised by surface interactions or some other well defined means, is the responsibility of a single specialised algorithm, whose individual results are combined to form the full global illumination solution.

Handling the wide variety of light transport features robustly poses a tough problem when designing global illumination algorithms. Chapter 4 discussed in detail the implementation and principles behind Monte Carlo path tracing and photon mapping approaches, along with an overview of their respective advantages and limitations. Monte Carlo path tracing and its derivatives have difficulty evaluating highly specular light paths, especially in conjunction with high frequency lighting features. Conversely, photon mapping techniques perform poorly for less deterministic paths where the underlying photon distribution is distorted by interactions with diffuse surfaces, and not well distributed with respect to the camera importance.

In this chapter, a new hybrid rendering approach is described that seeks to solve some of the problems currently associated with state-of-the-art light transport algorithms from these two paradigms. The approach presented here employs a classification strategy in path space during sample evaluation to classify paths into distinct subsets. These subsets can then be evaluated with specialised derivations of the Monte Carlo path tracing and progressive photon mapping algorithms, adopting the advantages of each paradigm while aiming to disinherit their weaknesses under particular classes of light transport.

First, the limitations and inefficiencies of existing unbiased Monte Carlo path tracing, photon mapping and their respective extensions are first discussed to motivate their use in a hybrid algorithm. Following this a discussion of previous works that leverage multiple paradigms to solve global illumination is presented, and some practical and theoretical implications for such techniques. The remainder of the chapter will detail the path space separation and the use and roles of path tracing and progressive photon mapping as an effective hybrid rendering technique. This chapter finishes with presentation of the results of this novel approach, showing its success alongside an in depth discussion and its potential as a platform for future research.

Contributions

The novel work presented in this chapter seeks to eliminate high frequency image noise through a combination of Monte Carlo path tracing and stochastic progressive photon mapping. The path tracing process is separated into path generation and radiance evaluation at each vertex, allowing radiance contributions to be selectively evaluated on the fly, depending on the path characteristics. A set of pattern matching filters over path space are defined, enabling each subpath contribution to be evaluated individually according to its surface interactions. Standard

unbiased path tracing is used to evaluate the largely non-specular paths. This takes advantage of the path tracer's ability to form explicit connections to light sources and provide the means for effective sample placement. An adaptation of the stochastic progressive photon mapping technique of Hachisuka et al.[HJ09] is employed to render fast, crisp caustics with low bias. Despite exhibiting bias early on in rendering, the desirable low memory and vanishing error properties of SPPM, which are commonly associated with unbiased rendering techniques, are leveraged and further refined. In doing so the following novel contributions are presented:

- Efficiently combining path tracing and progressive photon mapping techniques to produce a vanishing error, multi-pass progressive algorithm (referred to as PTPPM).
- A set of filtering criteria over path vertices are proposed, to reduce potentially high variance samples during Monte Carlo rendering using on the fly vertex contribution filtering.
- An efficient and progressive means to render unbiased non-caustic lighting and caustic lighting.
- Superior reduction in quantitative error, resulting in better convergence rates when compared to both standard path tracing and stochastic progressive photon mapping.

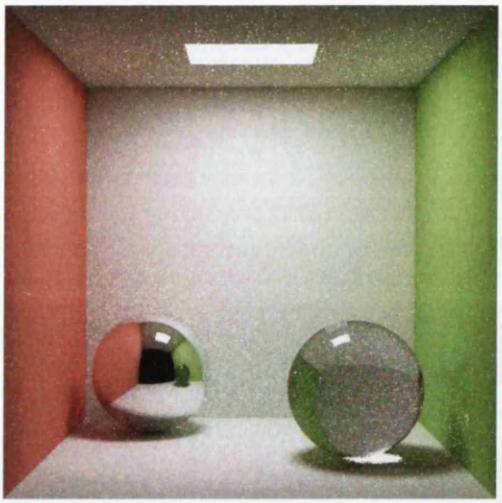
In turn these techniques give rise to a number of desirable properties for rendering complex scenes:

- Caustics and difficult specular path interactions are handled that are otherwise not possible using unbiased Monte Carlo methods
- Good sample distributions and variance reduction techniques are available for non-caustic lighting, utilising image space stratification during path tracing
- Reducing photon mapping to a sub-set of path space results in lower memory consumption and faster density estimates resulting in more efficient caustic evaluation.

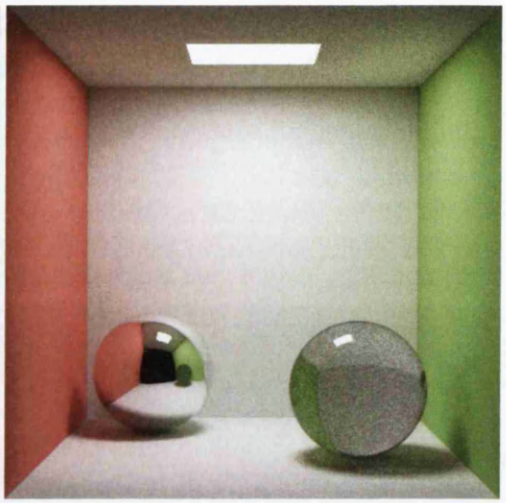
The basis of the work presented in this chapter has undergone peer review and has been published and presented at the *Computer Graphics International 2012* conference and was further selected to appear in a special issue of *The Visual Computer* journal [DJM12].

5.2 Motivation

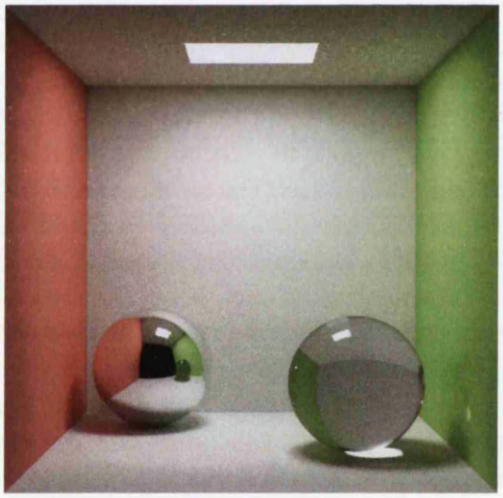
In this section, further motivation for the incorporation of path tracing and stochastic progressive photon mapping will be discussed, and the limitations of unbiased vertex based approaches and density estimation techniques will be presented along with the underlying causes which are addressed by the novel work.



(a) Full global illumination with 128 samples



(b) Samples from (a) with caustics omitted



(c) Full global illumination with 16x samples of (a)

Figure 5.1: *The addition of caustics results in increased sample variance for Monte Carlo estimators. (a) Full path tracing shows distracting high frequency noise. (b) By omitting the direct caustic paths, visual improvements are noticeable, but at the expense of losing important features. (c) The full global illumination solution with 2048 samples, where noise is still visible.*

5.2.1 Limitations of unbiased Monte Carlo methods

Unfortunately, path tracing suffers from spiked noise in the presence of highly peaked BRDFs and the generation of caustics in non-trivial cases. In these scenarios, the probability of sampling light sources through specific combinations of interactions is low, resulting in high sample variance. Even after evaluating many thousands of samples, the resulting variance can still show up as high frequency spiked noise in the image. High energy caustic lighting is a prime example exhibiting these difficulties, due to its impact on the final image. It can be seen that fil-

tering out the direct caustic components from the sample set provides a significant decrease in variance, at the expense of removing the important information and realism that caustic paths bring to such images (Figure 5.1). The implications of caustic lighting and specular materials in real world scenes make such paths a candidate for attention and improvement.

The fundamental problem that restricts current unbiased vertex based light transport techniques, is that the narrow BRDFs of specular materials inhibit the use of next event estimation and explicit vertex connections, decreasing the algorithm's effectiveness. Implicit connections are therefore relied on to produce radiance contributions, emphasising the use of effective sampling distributions. These problems are compounded by the use of BRDF importance sampling, since specular paths with high throughput can create unpredictable peaks in the importance sampling distribution.

This is increasingly the case for small light sources, common in scenes modelled on the real world, where paths with implicit connections are difficult to explore, leading to poor convergence. For point or directional light sources, useful for approximating small or distant light sources, direct caustics cannot be evaluated at all by camera paths since explicit sampling of the light source not possible from a specular surface, and implicit sampling is not possible for light sources with infinitesimal surface area.

The problem of implicit sampling is exacerbated when dealing with reflected or refracted caustics through specular-diffuse-specular (*SDS*) surface interactions. Such paths are important for rendering interior scenes (such as a room with windows illuminated from the outside) or scenes under artificial lighting (where emitted light must pass through the glass casing of a light bulb). Under these conditions, simulating *SDS* paths is essential, especially if such scenes contain specular objects, for example a mirror. Light must pass through a specular interface (the window or light bulb glass, *S*), before undergoing scattering by a diffuse object (the *D*) in order to be observed in the mirror (the final *S*). In *SDS* paths, the number of consecutive implicit connections required to connect the light source and camera is further increased, thus a small change in the outgoing direction at the enclosed diffuse vertex or the incident angle at a specular surface can lead to large changes in the resulting path geometry (Figures 5.2a to 5.2c). In the presence of high frequency features such as small light sources or hard shadows following the *SDS* interactions, variance of the pixel estimator can increase significantly.

In the simplest case, assume there exists a valid path of the form *ESDSL* that should be sampled by an estimator. Under path tracing, the only option is to extend the camera path at each vertex, using BRDF importance sampling to connect the camera and light source implicitly. In the bidirectional case, a light path *LSD* or camera path *ESD* can be generated, perhaps both with high density, but their vertices cannot be connected explicitly unless on convex surfaces. Thus the use of BDPT provides no additional benefit, falling back to a unidirectional path construction from the camera (as in Figure 5.2a).

Even with the use of Markov-chain Monte Carlo methods such as Metropolis light transport, small mutations in the random variables used to generate *SDS* paths can be ineffective or highly inefficient. Mutation cannot be performed at specular vertices due to their Dirac distribution, limiting the available mutations. Mutations around L in position or emission direction are ineffective since implicit connections to the camera are not supported (and impossible in the case of a pinhole camera model). Modifying the outgoing direction of the primary ray from E or the enclosed diffuse vertex D relies once again on an implicit connection to the light

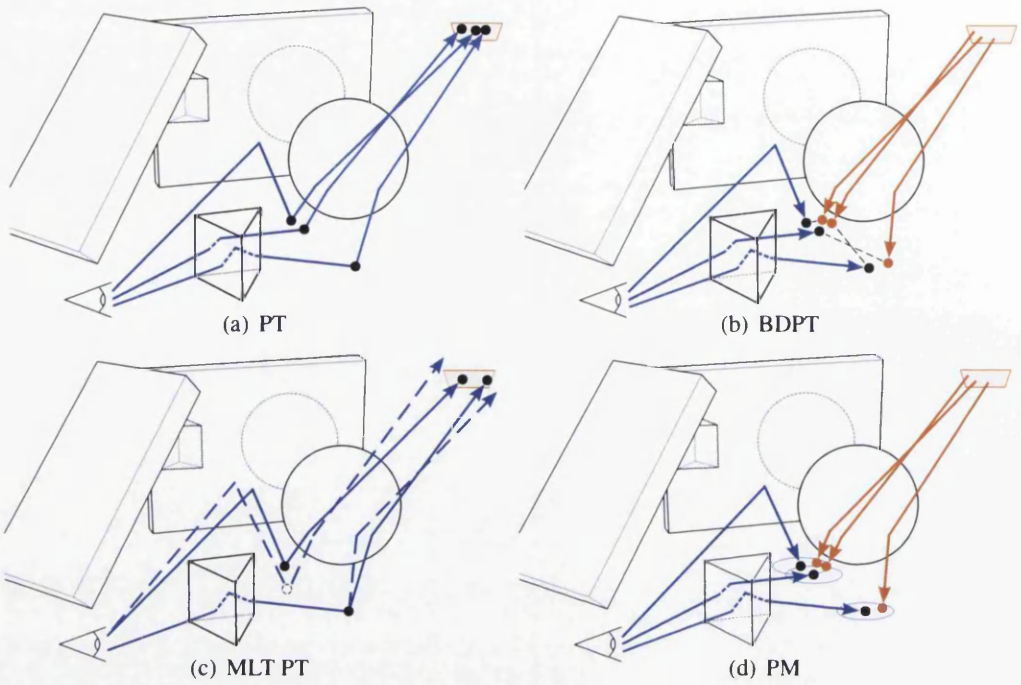


Figure 5.2: *Sampling a path of the form ESDSL can be problematic with (a) path tracing, (b) bidirectional path tracing and (c) Metropolis light transport. (d) Using density estimation with light sub-path caching allows multiple such paths to be evaluated efficiently by photon mapping algorithms.*

source, although this time more informed. To maintain an unbiased estimator this requires the change to be propagated along the path re-evaluating its terms including costly visibility queries, incurring the overhead of re-tracing a significant portion of the full path (Figure 5.2c). Whilst providing an advantage over standard MC methods, in the presence of small light sources mutation is also highly ineffective. Even small changes to the outgoing direction at E or D can change the path significantly due to the deterministic nature of specular reflections or refractions.

5.2.2 Photon Mapping for caustic illumination

The more relaxed constraints of photon mapping methods permit them to generalise visibility queries, allowing reduced variance at the cost of introducing bias. Using density estimation widens the path space encompassed by each sample, allowing better evaluation of paths containing complex specular interactions compared to exact point to point path space techniques. Taking the example of an *ESDSL* path, the *LSD* light sub-path and *ESD* camera path can be generated independently, and if they lie within the distance dictated by the gather kernel radius, are treated as connected. The amount of bias introduced is dependent on the distance between the end points of the paths in the domain of the filter kernel.

Unlike Markov-chain MC methods, the remaining sub-path vertices and E and L origins

are not modified, saving the computational overhead of additional ray tracing. Additionally, photons can contribute radiance to multiple camera paths, and each camera path can connect to multiple light sub-paths, efficiently reusing the illumination data to increase correlation and reduce high frequency noise in the image.

It has been established that classical photon mapping can have impractical memory requirements to provide high enough photon densities to reduce visual artefacts, and result in costly photon lookups.

Progressive photon mapping and its stochastic variant are effective in regions of high photon density such as caustics, and their ability to handle complex light paths has shown to superseed that of the original photon mapping, approach as discussed in Chapter 4. Decreasing the kernel radius as photons are collected reduces bias and introduces variance slowly in high density regions, since there are still a significant number of photons from which to draw a radiance estimate. Crucially, in such high density regions the initial gather radius of the kernel can be kept small without introducing high variance due to insufficient photons in the estimate.

In regions of lower photon density, (S)PPM exhibits poor convergence as the initial kernel radius may not provide sufficiently reliable radiance data to offset the increasing variance brought about by the poor photon density. Thus larger kernels are needed to keep variance low, in turn increasing the amount of initial bias in the estimate. As the final radiance estimate is a weighted average of all previous PPM iterations, bias introduced in initial passes is highly pervasive and still affects pixel estimates even after a considerable number of passes. This is a side effect of the progressive kernel bandwidth reduction. In practice, as the gather radii become small, it is increasingly likely that no or few photons will be collected for a given iteration i, and hence the unnormalised flux τ for the subsequent iteration $i+1$ will remain constant when $M_i(S)$ is zero or $M_i(S) << N_i(S)$:

$$\begin{aligned}\tau_{i+1}(S,\omega) &= (\tau_i(S,\omega) + \Phi_i(x_i,\omega))\frac{N_i(S) + \alpha M_i(x_i)}{N_i(S) + M_i(x_i)} \\ &\approx \tau_i(S,\omega)\end{aligned} \qquad \text{(4.10 revisited)}$$

Apart from progressively reducing per sample efficiency, this also means that more iterations are required in low photon density regions before the initial bias is eliminated by those iterations where $\Phi_i(x_i,\omega)$ is non-zero. For high density regions such as caustics the probability of non-zero radiance estimates is much lower, and hence even though the initial bias may be high, the probability of obtaining zero radiance estimates is reduced. In the limit, as the radius $R_i(S)$ becomes infinitesimal, $M_i(S)$ will inevitably reach zero, since the photon density is finite. Bias will therefore always be present in PPM in practice, though with negligible impact on the image for sufficient i.

It has been established in Chapter 3 that good sample distributions are key in the success of all Monte Carlo and quasi-Monte Carlo methods. Utilising such sampling sequences and patterns for photon sampling results in a set of primary photons that are well distributed with respect to the light sources in the scene. As photons are emitted, the sampling sequences aim to ensure that photons sharing similar points of origin on a light source, are propagated widely across the visible geometry. For subsequent interactions with the scene, the ability to control the underlying global distribution of these secondary photons is decreased. For diffuse BRDFs,

where the importance sampling functions are wide, the global photon distribution becomes analogous to that of multiple independent low density distributions each from a different emitter. Relying on light source and flux based sample distributions alone can therefore produce poor photon maps, exhibiting highly erratic densities and a distribution that does not follow the features of the illumination. The inclusion of glossy and specular interactions with narrow BRDFs can further enhance this distortion, where small changes in the random variables used for path construction cause large variations in the path geometry and thus the locations of the photons.

Furthermore, the resulting photon map is often poorly distributed with respect to camera importance. Even assuming an optimal photon distribution, using a perspective camera model means that visible regions of the scene will likely be under or over sampled, reducing efficiency or increasing variance, bias or both. The angle of incidence of primary camera rays with the surface can also play a major part in the resulting image quality, noticeable in the images rendered with SPPM later in this chapter. Monte Carlo methods do not suffer this problem as illumination arriving at diffuse surfaces is a result of sampling according to both the camera importance and explicitly from light sources.

The slow convergence properties of PPM can be a drawback of relying solely on density estimation for radiance estimates. Compared to the $O\left(N^{1/2}\right)$ convergence of unbiased MC methods, PPM exhibits poorer convergence of $O\left(N^{\alpha}\right)$, reliant on the radius reduction parameter $0 < \alpha < 1$. Optimal values of α have shown to be around $\frac{2}{3}$ [HPJ12] [KZ11]. Modifying α to match the asymptotic convergence of path tracing methods is impractical and leads to poor efficiency. If α is too small, the gather radius is reduced such that the rate at which variance is introduced exceeds that of the bias reduction brought about by averaging successive radiance estimates, increasing the total amount of error in the estimator.

5.2.3 Hybrid approaches

Hybrid approaches to global illumination have been explored previously. Wallace et al.[WCG87] and Sillion and Puech [SP89] extended finite element radiosity to incorporate specular reflections and refraction by the inclusion of an additional ray tracing pass. Shirley [Shi90] separated the direct lighting computation from the radiosity pass in order to improve hard shadows and account for bump maps. More complexity was integrated into such a system by Chen et al.[CRMT91], who rendered caustics using a light tracing pass, though without the ability to handle reflected and refracted caustics.

Suykens and Willems explored multi-pass approaches for combining (bidirectional) path tracing and radiosity methods for global illumination [SW99, Suy02]. In their proposed system, bidirectional path tracing is utilised as a final pass to account for all light transport not otherwise handled in preceding passes, ensuring full coverage of the path space.

The reliance on finite element methods makes these techniques poorly suited to high frequency lighting and geometry without finer or adaptive subdivision of the patches, increasing computational or storage requirements. The added complexity of progressive variants, which require tracking residual illumination that is not accounted for, also makes them less favourable for many present day applications.

Recently Dammertz et al.[DKL10] introduced a biased progressive method, which also divides path space to allow specialised algorithms to compute a solution for each subspace, in a similar fashion to the work presented here. They separate paths into three categories; indirect illumination EDL and $EDD(S|D)^*L$, caustics $EDS(S|D)^*L$, and specularly gathered illumination. Utilising virtual point lights (VPLs) provides fast convergence and high coherence between pixels reducing noise for the indirect illumination. Virtual point light based methods have difficulties when incorporating complex scattering models such as subsurface scattering and volumetric effects since eye paths are terminated at the first diffuse hit point and rely on point to point gathering. Light propagating through non-specular materials therefore cannot be easily accounted for, requiring special handling via local density estimation (as is used in photon mapping) or continuation of the camera path which will lead to the high frequency noise typical of Monte Carlo methods; absence of which is the main benefit of VPL approaches.

In the rendering framework proposed by Dammertz et al., caustics are collected using a histogram (or binning) approach, gathering photons into fixed sized kernels based on ray differentials at the first diffuse camera vertex. Using fixed size bins poses similar limitations to those found in classical photon mapping, where boundary, proximity and topology bias are easily introduced in the presence of complex geometry and illumination discontinuities. Additionally, it is not clear how non-spherical or arbitrarily shaped kernels could be employed to help reduce proximity bias such as light leaking that lead to visual artefacts. The requirement for pixel sub-sampling for anti-aliasing in addition to these fixed size bins results in photons being collected that are outside the pixel footprint, which can cause noticeable blurring for sharp caustics and paths undergoing complex refractions. Although stochastic progressive photon mapping also exhibits similar behaviour early on with respect to kernel size, the gradual radius reduction ensures this bias vanishes with increased passes, and complex caustics are rendered faithfully.

5.2.3.1 Drawbacks of hybrid approaches

Incorporating two or more algorithms into a single framework also has its disadvantages. As they subsume existing approaches, the task of implementing a hybrid algorithm can be complex, requiring an understanding of the existing algorithms and their execution flow. This becomes increasingly more important when dealing with highly parallel architectures such as the latest multi-core processors and general purpose GPU hardware. Understanding the underlying performance bottlenecks and memory access patterns of an algorithm is essential in order to efficiently incorporate it within a hybrid technique. In addition to access patterns, memory consumption can also become problematic since data required under each original method is often still necessary, at least in part.

Furthermore, it is important to ensure that those properties unique to each algorithm that make them desirable are not lost upon forming a hybrid approach. Practical properties such as progressive estimation of the integral, consistency, systematic bias and view dependence need to be taken into account and preserved where possible.

The two pass approach taken by classical photon mapping does not fit well with the progressive nature of unbiased Monte Carlo methods. Both the view independence of photon mapping and the progressive characteristics of MC methods would be lost in a hybrid approach,

decreasing the usefulness of such a technique. Thus combining a progressive photon mapping approach with unbiased Monte Carlo path tracing retains the usefulness of a progressive renderer.

In recent years, hybrid approaches have seen less attention due to the ability of modern techniques to handle complex geometry, materials and light transport using a single algorithm. Unbiased path tracing and biased photon mapping are some such techniques which, with their respective extensions and improvements, have been used as the benchmark for state of the art research for a number of years. However, their limitations and inefficiencies discussed above and in the previous chapter remain prevalent, and the benefits brought about by hybrid techniques are still applicable.

In this work, the properties of progressive photon mapping are leveraged to provide better evaluation of caustic lighting and complex *SDS* paths, combined with path tracing techniques whose strengths lie in solving diffuse and glossy direct and indirect lighting under complex conditions, with good convergence characteristics. This results in a hybrid progressive algorithm exhibiting reduced variance and error characteristics with vanishing bias. Additionally, the progressive nature of the two techniques is preserved, and the slower convergence of photon mapping methods for low density regions is offset using path tracing, whilst still taking advantage of the effectiveness of photon tracing techniques in areas of high photon density.

5.3 Overview

The hybrid algorithm presented in this chapter operates in a multi-pass progressive manner, focussed on providing high quality results and efficient convergence characteristics for diffuse lighting and caustic lighting. The first pass traces paths from the eye using standard path generation techniques discussed in Chapter 4, including BRDF importance sampling and Russian roulette path termination. Unlike path tracing, radiance contributions at each vertex are postponed until after the full path is generated. Instead, the surface interactions of each vertex are used to classify the associated sub-paths into either the diffuse or caustic subspace. Radiance contributions are calculated for diffuse sub-paths and added to the image buffer. Contributions to the caustic subspace are skipped at this stage, and evaluated in a second pass using a modified PPM technique.

Caustic photons are generated at the light sources and traced through the scene, deposited on diffuse surfaces. During this photon path generation, the path signature is built up on the fly, and the path is terminated prematurely if the signature does not match the pattern for the caustic subspace, or else terminated using Russian roulette. A kD-Tree is constructed around the deposited caustic photons in order to accelerate photon gathering. Hitpoints from the initial camera path generation are utilised to gather photon flux and compute caustic radiance estimates for each pixel, after which the photons are discarded. After each iteration, the intermediate radiance estimates computed by the two algorithms can be combined to produce a full global illumination solution. Maintaining a disjoint path space for each algorithm means that no bias is introduced at the compositing stage, and without the need for complex or expensive weighting functions. The core components of the two passes are summarised in Algorithms 2 and 3 respectively, and details will be explained in the remainder of the chapter.

Algorithm 2 Diffuse lighting pass via path tracing

1: **for all** pixels P in image **do**
2: $P(x_d) \leftarrow null$
3: $x_{i-1} \leftarrow x_0$ ▷ Generate point on the camera x_0
4: **while** path not terminated **do**
5: $x_{i+1} \leftarrow r(x_i, \omega_i')$
6: $\bar{x}_{i+1} \leftarrow \bar{x}_i + x_{i+1}$ ▷ Extend x_i using $f_r(x_i, \omega_i, \omega_i')$ and $r(x_i, \omega_i')$
7: **if** $P(x_d) = null$ and $f_r(x_i)$ has diffuse component(s) **then**
8: $P(x_d) \leftarrow x_{i+1}$ ▷ Store first diffuse hitpoint, for reuse in Algo. 3
9: **for all** vertices x_i in $\bar{x}_n$ **do**
10: **if** x_i contribution(s) are "diffuse" **then** ▷ Apply path space filtering
11: $s \leftarrow$ Generate direct lighting samples
12: $L_{direct}(x_i) \leftarrow$ Evaluate direct lighting at x_i using s
13: $L_d \leftarrow L_d + L_{direct}(x_i) \cdot T(\bar{x}_i) \cdot f_r(x_i, x_{i-1} \leftarrow x_i, x_i \rightarrow x_{i+1})$
▷ Update diffuse radiance L_d

Algorithm 3 Caustic illumination pass using photon tracing

1: **for** $j \leftarrow 0 \rightarrow n$ photons **do**
2: $y_j \leftarrow r(y_0, \omega_0)$ ▷ Generate primary photon from emitter at y_0
3: **while** photon path not terminated **do**
4: $\bar{y}_{j+1} \leftarrow \bar{y}_j + x_{j+1}$ ▷ Extend y_j using $f_r(y_j, \omega', \omega)$ and $r(y_j, \omega_j)$
5: **if** photon contribution of y_j is "caustic" **then** ▷ Apply path space filtering
6: Photon map $\leftarrow y_j$
7: Build kD-Tree around photon map
8: **for all** pixels P in image **do**
9: Retrieve $P(x_d)$ from Algo. 2 Line 3
10: Gather N nearest photons around $P(x_d)$
11: $L_c \leftarrow L_c +$ photon flux of the N photons
12: $L \leftarrow L_d + L_c / n$ ▷ Combine the diffuse L_d and normalised caustic L_c components

5.4 Path space separation and filtering

The regular expression and path notation introduced by Heckbert [Hec90] is used once again in order to concisely describe the seperation of path space. Pattern matching can be applied over these regular expressions in order to selectively evaluate the paths generated by both the unbiased Monte Carlo and photon mapping techniques. First, let us construct an expression that encompasses all paths in our caustic subspace:

$$E(S|D)^* D^+ S^+ D^? L \tag{5.1}$$

Notice firstly that this incorporates both direct and single-bounce indirect caustics due to the optional diffuse vertex, $D^?$. Indirect caustics are also a common source of high variance in path tracing, even viewed through multiple diffuse bounces. Low probability paths generated by

inaccurate BRDF importance sampling at the diffuse vertex, D^+, coupled with high luminance caustic paths created by the S^+L sub-path can vastly increase variance. Even when viewed indirectly through other diffuse vertices, the unbounded variance of such interactions can create high intensity noise in the pixel estimator, more commonly when the distance between surfaces is small and even diffuse BRDFs do not dissipate the high radiance sufficiently to reduce variance.

Based on this caustic regular expression the two path subspaces can be defined. The camera generated path space, evaluated using path tracing, encompasses primarily diffuse paths with higher path probabilities due to explicit light source connections, resulting in lower variance:

(D1) ES^*L: Light sources viewed directly or indirectly via specular surfaces,

(D2) ES^*DL: Direct lighting (optionally viewed via specular surfaces),

(D3) $E(S|D)^*DDL$: Multiple bounce indirect diffuse lighting viewed via other surface interactions.

Although paths with diffuse-specular interactions are included in this subspace, potentially leading to high pixel variance, the successive diffuse vertices (*...DDL*) mean that these paths are of relatively low luminance. Additionally, evaluation of single and multiple bounce diffuse lighting is also performed at this stage as it responds well to sample stratification. The lower frequency BRDFs present non-deterministic vertex interactions meaning the well distributed samples are more effective at reducing estimator variance. Using explicit light source sampling allows the application of multiple importance sampling for direct lighting at diffuse vertices, which is not applicable when handling direct lighting with density estimation. The remainder of the path space is handled by an adaptation of progressive photon mapping, relying on photons distributed from the light sources:

(C1) $LS^+D(S|D)^*E$: Direct caustics, formed from the light source.

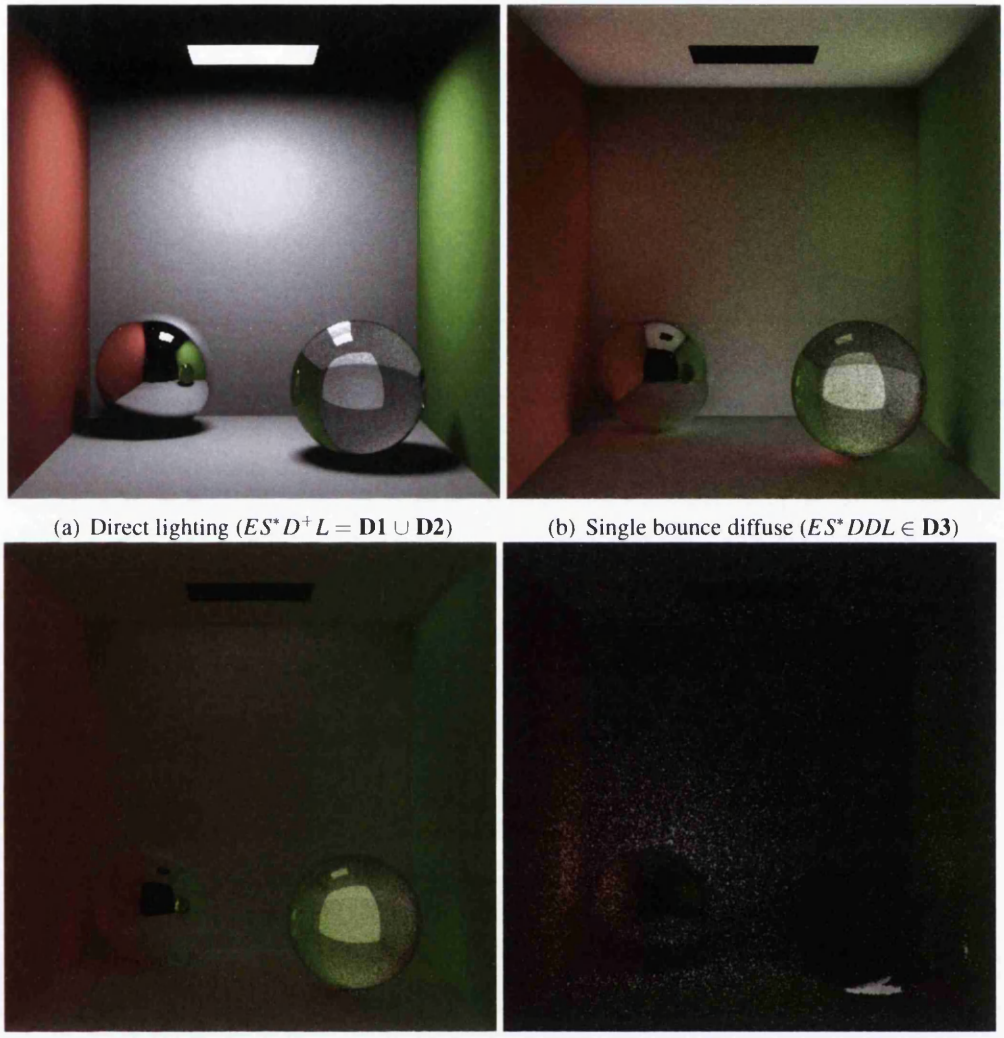
(C2) $LDS^+D(S|D)^*E$: Indirect caustics, reflected from a single diffuse surface.

For simplicity the path tracing subspace (**D1** to **D3**) will be referred to as "non-caustic" or "diffuse" and the path subspace handled by the progressive photon mapping algorithm (**C1** and **C2**) as "caustic". All possible paths are included in these two subspaces, allowing the evaluation of the full global illumination solution, where the two techniques have no overlapping path space.

5.5 Path tracing

The first pass in each iteration evaluates the radiance obtained from the non-caustic path subspace using Monte Carlo path tracing. In the implementation proposed here, the path generation and radiance estimation steps are isolated to allow for inclusion of the vertex classification and selective radiance evaluation.

During path generation, the random variables required to construct the camera path itself are generated and the path traced through the scene. The geometric and surface interaction data



(a) Direct lighting ($ES^*D^+L = \mathbf{D1} \cup \mathbf{D2}$) (b) Single bounce diffuse ($ES^*DDL \in \mathbf{D3}$)

(c) Multiple bounce diffuse ($\mathbf{D3} \setminus ES^*DDL$) (d) Caustic and indirect caustic ($\mathbf{C1} \cup \mathbf{C2}$)

Figure 5.3: *Filtering over path traced samples allows noisy direct and indirect caustics (d) to be separated from lower variance samples (a)-(c) and more efficiently generated via photon tracing.*

necessary to allow radiance calculations to be postponed is recorded for each path vertex x. This includes the path throughput up to the current vertex, $T(\bar{x}_i)$ and reference to the surface's BRDF $f_r(x, \omega, \omega')$:

```
PathVertex{
  Position x              World space position
  BRDF* f_r               Reference to BRDF
  Colour T(x̄_i)           Path throughput from x_0 → x_i
}
```

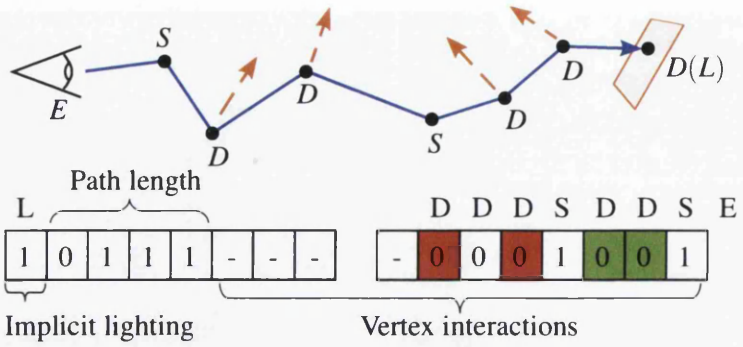


Figure 5.4: *Optional implementation detail of an example path and its corresponding bit string. The lowest 11 bits store a binary representation of the path vertex interactions, 4 bits are reserved for the path length and the topmost bit indicates an implicit light source connection.*

To provide a concise means to classify path contributions, a bit-string can be computed during path generation to encodes the path's interactions with the scene (Figure 5.4). The state of each bit of the path vertex representation indicates whether the path underwent a specular (cleared bit) or non-specular interaction (set bit). This bit-string is an optional implementation detail, since memory usage is not a concern, and other representations are possible.

In order to improve efficiency and ensure path termination in specular environments (where infinite reflection can occur), renderers often limit the maximum allowed path length. In this implementation a six vertex limit is chosen, which is long enough for most paths to terminate stochastically without any bias. In the rare cases where paths reach this limit, the path throughput is typically small enough that any bias is minimal.

The path length is encoded using 4 bits, to ensure that the correct bits are used in the pattern matching computation. A single bit is used to denote whether or not the final path vertex implicitly hits a light source. This is useful when identifying implicit caustic paths generated via path tracing, assisting in the lazy evaluation of direct lighting samples for the path's final vertex.

Pattern matching is applied to this bit string using **C1** and **C2**, to filter out caustic contributions. Operating over compact bit strings means that pattern matching is fast, relying only on cheap integer comparisons and bitwise operations in order to match path signatures with regular expressions. To simplify the regular expression computation, and as the specular and non-specular nature of the path vertices maps naturally to binary numbers, compact lookup tables can be pre-computed to match common vertex patterns and sub-paths.

To account for the vertex classification step, the radiance evaluation for each vertex contribution is delayed until after path generation. The advantageous side affects of this are two-fold. First, lazy evaluation can be applied for all light source sampling. Normally, during path generation the number of random variables required for lighting computations is unknown. Both the inclusion of specular vertices, and the implicit connection of the path with light sources can reduce the need for random variable generation. After the complete path is known, the sampling requirements are also known precisely, based on the BRDF sampled at each path vertex, and the classification given to each vertex of the path which may reduce the need for

direct lighting. Second, the effect of this lazy evaluation permits the distribution of samples to be potentially improved, as the exact number of samples and their relationship is known and can be generated together.

Generation of the necessary direct lighting samples is performed lazily, reducing the overhead of generating wasted samples. The generation of such samples is skipped when the path vertex is:

- Sampled using a specular BRDF, where direct lighting is not possible.
- Diffuse, and precedes a specular interaction and is therefore part of a DS^*L caustic,
- Diffuse and precedes a specular-diffuse interaction, and thus an indirect DS^*DL caustic path,
- Located on a light source, as light sources themselves are non-scattering,

In the case of diffuse vertices that precede an implicit light source connection, a direct lighting sample is still generated since multiple importance sampling can be applied to potentially reduce variance. The implicit ray connection to the light source has already been checked for visibility during path construction, so utilising multiple importance sampling for variance reduction can have minimal costs in such cases, if the relatively cheap computation of the MIS weights is calculated prior to checking visibility.

During path classification, it must also be ensured that implicit lighting from emissive surfaces is omitted after diffuse-specular interactions (...DS^*L contributions) as this corresponds to direct caustic lighting. As the ray generation for such implicit caustics are generated anyway, these contributions could also be added to an additional unbiased caustic buffer with minimal overhead. This could be combined with the caustic buffer produced by the progressive photon mapping technique developed later, improving convergence later in the rendering. Coming up with a weighting operation to combine these buffers is left for future work, but could be useful in reducing the bias present in the photon mapping estimator and improving the overall convergence of the caustic lighting.

After lazy generation of the necessary samples, the direct lighting is computed at each path vertex using the BRDF and geometry information stored at each vertex, and the final set of radiance values is added to the unbiased diffuse image buffer.

Using unbiased Monte Carlo methods provides accurate and robust evaluation of the non-caustic illumination, whilst eliminating the noise that is commonly introduced by caustic paths, which are dealt with in a second pass using density estimation.

5.6 Caustic evaluation

The second pass of each iteration evaluates caustic lighting using a derivation of progressive photon mapping. Having generated a set of eye paths to evaluate the diffuse sub-space, vertex data from the first pass can be reused to identify visible regions from which photon density estimates can be performed. Similarly to the path tracing pass, only a subset of the complete

Scene		Caustic Paths Direct	Caustic Paths Indirect	Photon Mapping Deposited	Photon Mapping Emitted	Photon Mapping % deposited
Cornell Box	*PT*	153k	218k	-	-	-
	PTPPM	(122k)	(172k)	4.71mil	52.2mil	9.0%
Ring	*PT*	1328	12.8k	-	- -	
	PTPPM	(1069)	(10.1k)	5.40mil	37.7mil	14.3 %
Shapes	*PT*	16.7k	10.0k	-	-	-
	PTPPM	(12.3k)	(7.31k)	6.22mil	81.1mil	7.66%

Table 5.1: *Comparison of the number of caustic paths generated using path tracing and the hybrid method proposed here, know as PTPPM. The caustic photon mapping pass evaluates many times more caustic samples than using path tracing alone. The first two columns for PTPPM (in brackets) show the number of radiance contributions excluded by vertex contribution filtering. Both algorithms were run for 5 minutes on each scene.*

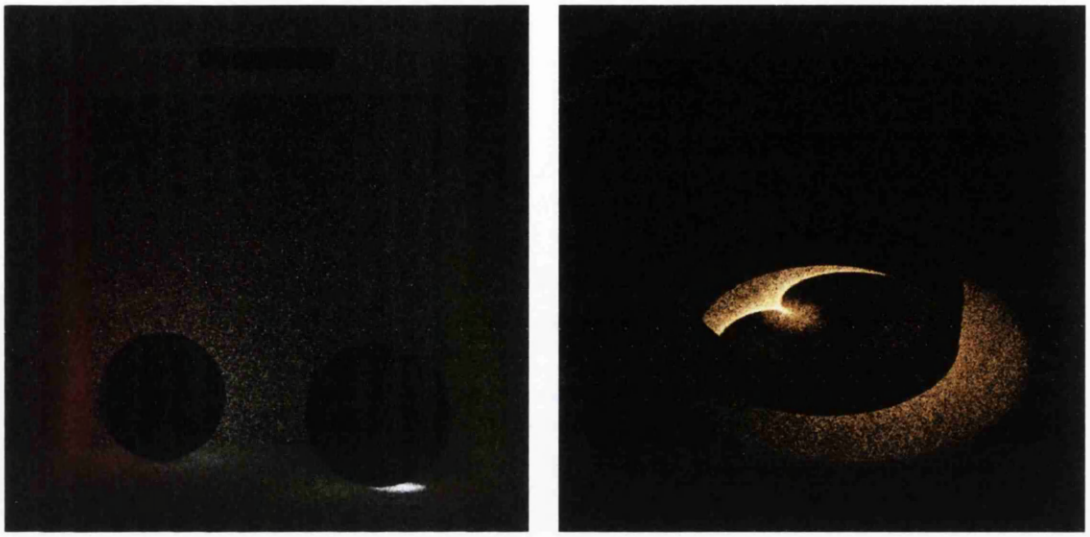
path space needs to be sampled, reducing the computational and memory costs in comparison to full photon mapping approaches.

During the initial iteration of the algorithm, the local statistics need to be initialised to allow photon gathering in keeping with the original formulation of PPM [HOJ08]. During the preceding path tracing pass, the first hitpoint with a non-specular BRDF component is found for each camera path. The local statistics for the progressive photon mapping pass are initialised, at these primary vertices. The pixel footprint described by the ray differentials at the primary vertex is used to dictate the initial kernel radius. This ensures that even in regions of low photon density the initial bias introduced into the progressive density estimate is kept to a minimum.

For subsequent iterations the primary hitpoints are refreshed from the preceding path tracing pass, providing the points x_i within the pixel region S around which photons are gathered. Often it is necessary to generate multiple camera paths per pixel in Monte Carlo rendering, to perform anti-aliasing and estimate distributed effects, as well as improved direct lighting estimates. This is exploited in this work to enable improved initial radii values for pixel statistics by averaging over the ray footprints of multiple camera hitpoints. This is especially helpful in the presence of geometric discontinuities and distorted or highly anisotropic ray footprints caused by glossy or specular reflection.

5.6.1 Photon tracing

Having found primary hitpoints for the current iteration, photons need to be traced to obtain a coarse, discrete approximation to the caustic illumination in the scene. As with the path tracing phase, only the subset of all possible paths are of interest. In similar fashion to the first camera path pass, pattern matching is applied on the fly during photon generation. As the path space of caustics is more restrictive, the termination of a photon path is simpler, being terminated earlier when a photon has undergone multiple diffuse bounces, with vertices matching *LDD*.



Cornell Box (after 16 iterations) Metallic Ring (after 128 iterations)

Figure 5.5: *Visualisation of the cumulative caustic photons deposited during iterations of the photon tracing pass. Depositing only direct and indirect caustic photons changes global uniformity of the photon map significantly, and reduces the number of photons in the map to around 9% and 14% for the Cornell Box and Metallic Ring scenes respectively. This improves the performance of the photon gathering pass, offsetting the additional time spent computing the diffuse lighting using unbiased techniques.*

For scenes with little or difficult caustic lighting, the number of emitted caustic photons in relation to non-caustic photons may be low. This improves memory consumption compared to using (S)PPM for the entire path space, and more efficient density estimation. Figure 5.5 shows a typical caustic photon distribution for two example scenes. The caustic photons are concentrated around the specular objects and provide a highly non-uniform global distribution.

The photon tracing stage is complete when the desired number of caustic photons have been deposited in the scene. For scenes with few or difficult to find caustic paths this can lead to the excessive rejection of photons, providing poor performance. Hence the photon tracing procedure is terminated after a large number of emissions if the desired caustic photons have not been stored. This improves efficiency in the case of difficult caustic illumination.

By storing an approximation of only the caustic lighting, the number of photons that need to be stored per pass can be reduced without loss of quality in the flux estimate. Caustic lighting is generally a focussing phenomenon which results in highly non-uniform densities throughout the scene. Contrarily, diffuse illumination has a more even global distribution due to the wide BRDFs of diffuse materials. Eliminating non-caustic photons thus increases the variance in density across the scene. This emphasises the need for adaptive spatial subdivision data structures to improve the performance of photon density estimates over highly non-uniform arrangements of photons. Large areas of sparse photons can be queried cheaply, whilst high density regions of concentrated caustic photons can be subdivided by the kD-Tree to reduce the number of distance queries required. For each iteration i, a kD-Tree is constructed around

the current set of caustic photons in preparation for photon gathering. Additionally, due to the reduced number of photons stored compared to (S)PPM, kD-Tree construction times are reduced in addition to the query times for each iteration (the results in Section 5.7 will provide further discussion). Once the caustic photons have been gathered for each pixel, the kD-Tree and the photon map is discarded and the photons are no longer needed.

5.6.2 Photon gathering

The photon gathering phase closely resembles that of the original (S)PPM method (recall Section 4.2.6). In contrast, only a subset of the complete path space is computed, so only partial radiance estimates for the caustic path space are required.

In similar fashion to the original, a radiance estimate is obtained and gradually refined by utilising the pixel hitpoints and the corresponding statistics. The updating procedure for the statistics remains the same, but instead operates over the caustic photon subset to compute the new accumulated photon count for caustics $N^c_{i+1}(S)$ using the number of caustic photons in the current gather kernel, $M^c_i(x_i)$:

$$N^c_{i+1}(S) = N^c_i(S) + \alpha M^c_i(x_i) \tag{5.2}$$

and the radius is similarly reduced using the global parameter α:

$$R_{i+1}(S) = R_i(S)\sqrt{\frac{N^c_i(S) + \alpha M^c_i(x_i)}{N^c_i(S) + M^c_i(x_i)}}. \tag{5.3}$$

Omitting the generation of non-caustic photons, does not affect the stochastically generated nature of the photon map. As a result, the photon distribution across a local region remains statistically constant between iterations; the same assumption previously imposed by (S)PPM.

As with the distributed ray tracing incorporated into SPPM, the path tracing pass of PTPPM generates new hitpoints for each pixel at each iteration. The $M^c_i(x_i)$ photons are accumulated over a region centred around this new hitpoint $x_i \in S$. The total unnormalised flux that these new caustic photons in x_i contribute to the radiance leaving S towards the camera can therefore be computed by:

$$\Phi^c_i(x_i, \omega) = \sum_{j=1}^{M^c_i(x_i)} f_r(x_i, \omega, \omega_j)\, \Phi_j(x_j, \omega_j). \tag{5.4}$$

The current caustic flux estimate for this region $\tau^c_i(S, \omega)$ can then be updated using the new flux estimate for this iteration, multiplying it by the change in radius from $R_i(S)$ to $R_{i+1}(S)$, to ensure an increasingly accurate estimate:

$$\tau^c_{i+1}(S, \omega) = \left(\tau^c_i(S, \omega) + \Phi^c_i(x_i, \omega)\right)\left(\frac{R_{i+1}(S)}{R_i(S)}\right)^2. \tag{5.5}$$

To produce the full global illumination image this partial flux estimate obtained via photon mapping needs to be normalised and combined with the path tracing radiance estimate for non-caustic lighting during image reconstruction.

Image Reconstruction After each iteration, an image can be produced by combining the pixel radiance estimates obtained from the two path subspaces. Since these subspaces evaluated by each method do not overlap, it is a trivial matter of summing the normalised radiance accumulated during path tracing, with the current progressive photon mapping estimate.

The flux estimates stored alongside the shared photon statistics first need to be normalised with respect to the total number of photons emitted so far across all iterations. Assume that all possible photon paths (both caustic and non-caustic) have been traced and separated to construct two photon maps. Using the shared radius, a number of caustic photons $M_i^c(x_i)$ and a number of non-caustic photons $M_i^p(x_i)$, could be gathered independently that lie within x_i. The contributing caustic and non-caustic flux values ($\Phi_i^c(x_i,\omega)$ and $Phi_i^p(x_i,\omega)$, respectively) can then be computed from each subspace independently. Given that both sets of photons are being gathered over the same region $R_i(S)$, the relative corrected flux values $\tau_i^c(S,\omega)$ and $\tau_i^p(S,\omega)$ remain proportional. After i iterations, $N_e^c(i)$ caustic photons and $N_e^p(i)$ non-caustic photons have been collected. As a result the radiance evaluation of the two path spaces can similarly be separated:

$$\begin{aligned} L(S,\omega) &= \lim_{i\to\infty} \frac{\tau_i(S,\omega)}{N_e(i)\pi R_i(S)^2} \\ &= \lim_{i\to\infty} \frac{\tau_i^c(S,\omega)}{(N_e^p(i)+N_e^c(i))\pi R_i(S)^2} + \frac{\tau_i^p(S,\omega)}{(N_e^p(i)+N_e^c(i))\pi R_i(S)^2} \end{aligned}$$

Radiance values for all non-caustic lighting are computed using the path tracing algorithm, hence the photon mapping non-caustic radiance estimate can be substituted with the unbiased path tracing estimate producing a full global illumination solution:

$$L(S,\omega) = \lim_{i\to\infty} \frac{\tau_i^c(S,\omega)}{(N_e^p(i)+N_e^c(i))\pi R_i(S)^2} + \int_S \int_{\Omega_x} f_r(x,\omega,\omega')L(x,\omega')(\mathbf{n}\cdot\omega')\mathrm{d}\omega'\mathrm{d}x$$

Thus obtaining an estimate for the radiance arriving at S scattered towards ω. To estimate the radiance at a particular pixel requires S to be the region of all surface points visible from the camera through the pixel, and Ω_x the area over the hemisphere at each location $x \in S$.

5.7 Results and discussion

The hybrid technique presented here has been tested on a number of scenes containing complex lighting effects with a particular focus on caustic lighting. Test scenes exhibit both reflected and refracted caustics, incorporating the specular-diffuse paths notoriously difficult in unbiased MC techniques. Direct and multiple bounce diffuse illumination is also prevalent, in addition to high frequency detail maps that provide improved visual fidelity.

All images are rendered on a PC with 8GB of memory and a 2.66Ghz Intel Core i7 CPU utilising 8 threads, providing more than sufficient memory and parallelism to provide a comparison of the different techniques on modern consumer hardware. All images are rendered at a resolution of 512x512 unless otherwise stated and a photographic tone reproduction operator (tone mapping) is applied following the work of Reinhard et al.[RSSF02] to improve clarity

and is responsible for some of the visual differences between the path tracing a photon mapping methods.

The path tracing implementation used for both the hybrid approach presented here and the standard path tracing use the same rendering framework for all aspects of image generation. The Monte Carlo path tracer uses variance reduction techniques described in Section 3.3, including next event estimation to evaluate direct lighting at each vertex, sampling from the light sources using a PDF based on their relative total flux. Russian roulette is introduced at the third path vertex using a threshold q based on the path throughput, or using $q = 0.5$ if the path throughput is higher. This ensures radiance estimates are made for the initial important bounces, and subsequently terminates paths with 50% probability along high contribution paths such as specular interactions to improve efficiency.

For each iteration of the hybrid approach, two camera paths are traced per pixel. The benefits of this are two-fold. First, because the pixel footprints are initialised using the ray differentials at the primary non-specular vertex, providing two ray footprints improves the initial kernel radius for pixels that cross geometric discontinuities and reduces the risk of artefacts that can be caused by kernel bandwidths derived from anisotropic ray footprints. Second, tracing an additional camera path better balances the convergence of the diffuse and caustic illumination since each density estimate typically evaluates multiple samples (one from each photon), but each MC path computes only one.

Comparisons with SPPM are also provided whose implementation follows that outlined by the original authors in [HJ09]. Parameters are also adopted from previous work, depositing 500k photons and using a radius reduction of $\alpha = 0.7$. A kD-Tree is built around the photon map in order to accelerate density estimation, using a median-split heuristic in order to provide a good trade off between construction time of the tree and photon gathering. Although heuristics such as the voxel-volume heuristic of Wald et al.[WGS04] have shown to be effective for classical photon mapping, their slow construction times are less suited to a progressive paradigm.

In the hybrid approach photon tracing for each iteration is restricted to depositing 50k caustic photons, or termination occurs after the emission of 400k photons. The initial criteria avoids the overhead of photon gathering becoming too large, when caustic illumination dominates the scene. The second limit is useful in scenes with difficult to sample caustics, where the time taken up by the photon tracing pass can become excessively long if left unbounded.

More optimal parameters can be chosen manually for each scene depending on the characteristics of the illumination, achieving a better balance between the refinement of caustic and diffuse lighting estimates. For scenes where caustic lighting dominates significant regions of the image, it can be favourable to increase the number of photons traced during each iteration to improve convergence for those regions.

The two passes of the algorithm are conducted sequentially for each iteration, and the multicore CPU is utilised by parallelising each pass. The generation of individual camera paths and each photon is independent, and thus parallelism can be conducted in the same way as for the two original algorithms. However, threads must be synchronised for photon gathering since either the complete photon map, or the complete set of pixel hitpoints for each iteration must be generated before caustic evaluation can be performed.

To evaluate the effectiveness of the proposed technique, both quantitative and visual com-

parisons are used. Equal time comparisons provide an indication of the relative efficiency of each technique to produce an accurate and perceptually correct image, compared to a reference render. Since all three algorithms discussed here are implemented in the same code base, developed as part of this thesis, the majority of their sub routines such as ray tracing, BRDF sampling and scene layouts are common to each, thus performance is most heavily influenced by the algorithm design itself. Cropped parts of the images are also shown to highlight where methods differ. The root mean square error (RMSE) is also computed during progressive rendering using a pixel-wise comparison to the reference solution. This provides a quantitative indication of the total numerical error across the image.

Using a hybrid method means that the number of samples evaluated for each path subset is reduced compared to the constituent algorithms. Despite this, regions of diffuse lighting have not noticeably suffered due to undersampling, due to the elimination of high frequency noise thanks to the separation of the vertex contributions. In comparison to SPPM, the new hybrid algorithm exhibits far less noise in diffusely lit regions, whilst producing caustics of equivalent visual quality across our test scenes.

5.7.1 Cornell Box

The first test scene (Figures 5.6 and 5.7) is inspired by popular modifications to the cornell box scene, including glass and chrome spheres lit from above by a moderately sized area light source. Spherical geometry with refractive and reflective properties ensures the formation of both focussed caustics on the floor, and highly dispersed caustic lighting is visible on the surrounding ceiling and walls. The ceiling and regions under the spheres are lit indirectly, with both hard and soft gradients visible around the spheres and the top of the walls. This a good test scene as it exhibits a number of common and varied phenomena.

Using standard path tracing, high frequency noise in the image is very apparent, obscuring the finer detail and fails to express the smoothness of the true illumination. Caustics are present but contain significant noise especially around the edges and in the multiple bounce caustic on the right hand side, refracted from the mirrored sphere through the glass.

Rendering with SPPM alleviates the high intensity spiked noise due to the use of sample correlation and photon caching providing far improved caustic illumination. However, parts of the image dominated by diffuse illumination contain noise caused by low density and high initial variance. Additionally, bias is introduced around the edges of the box, particularly noticeable in the shadowed regions that cross geometric edges. Due to the slower iterations of the SPPM algorithm (see Table 5.2), camera centric effects such as the reflection and refraction of the glass sphere take longer to resolve, despite the reuse of the photon map data.

Using the hybrid method, the otherwise objectionable high frequency noise caused by path tracing in conjunction with the specular convex geometry is removed, presenting a cleaner and less visually disturbing image. Caustic detail is provided by the photon mapping pass, improving in particular the multiple bounce caustics and their edges where such paths have low density when sampled from the camera. Boundary and proximity bias for illumination and geometric edges has been eliminated through the use of explicit point-to-point connections.

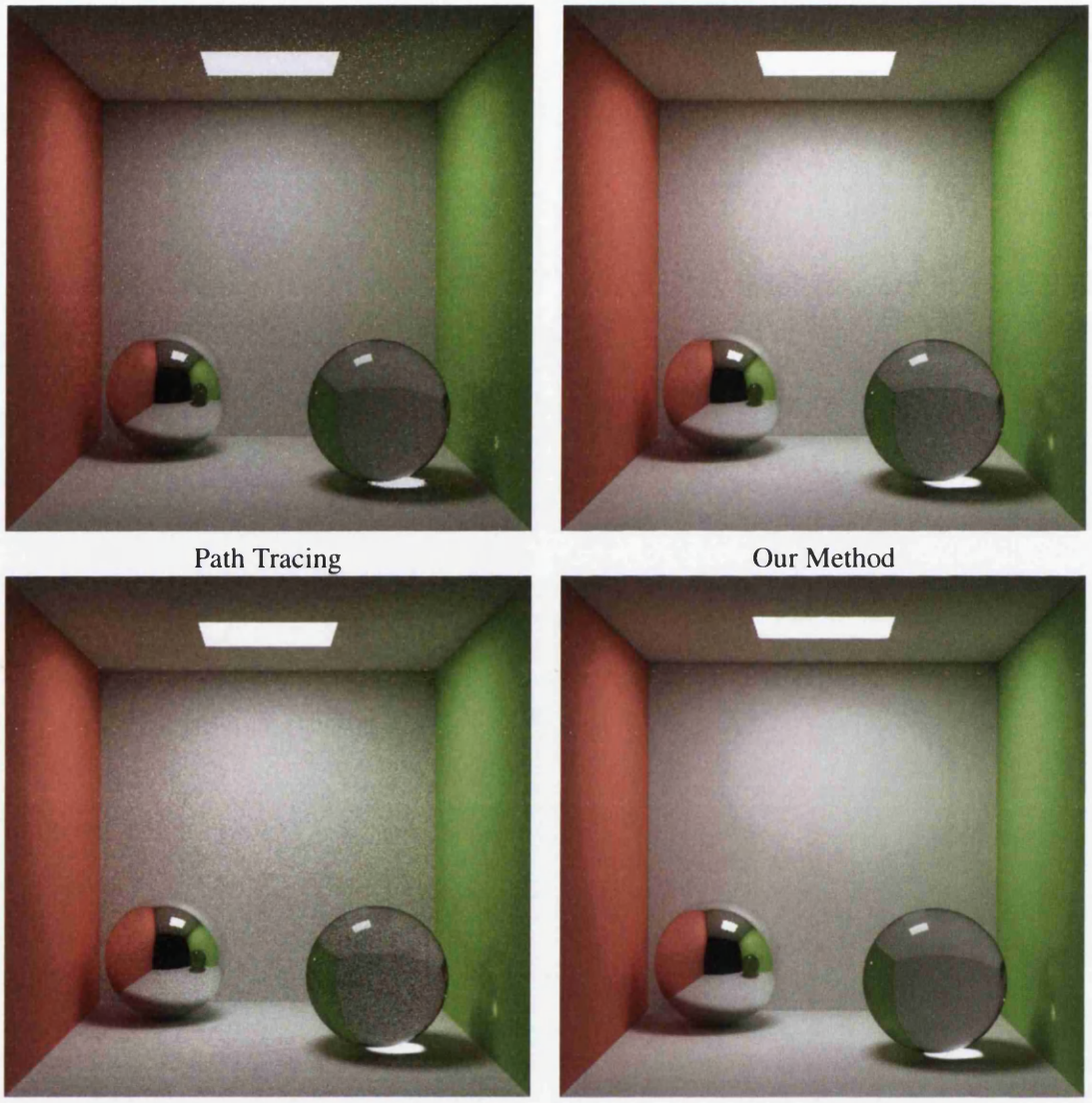
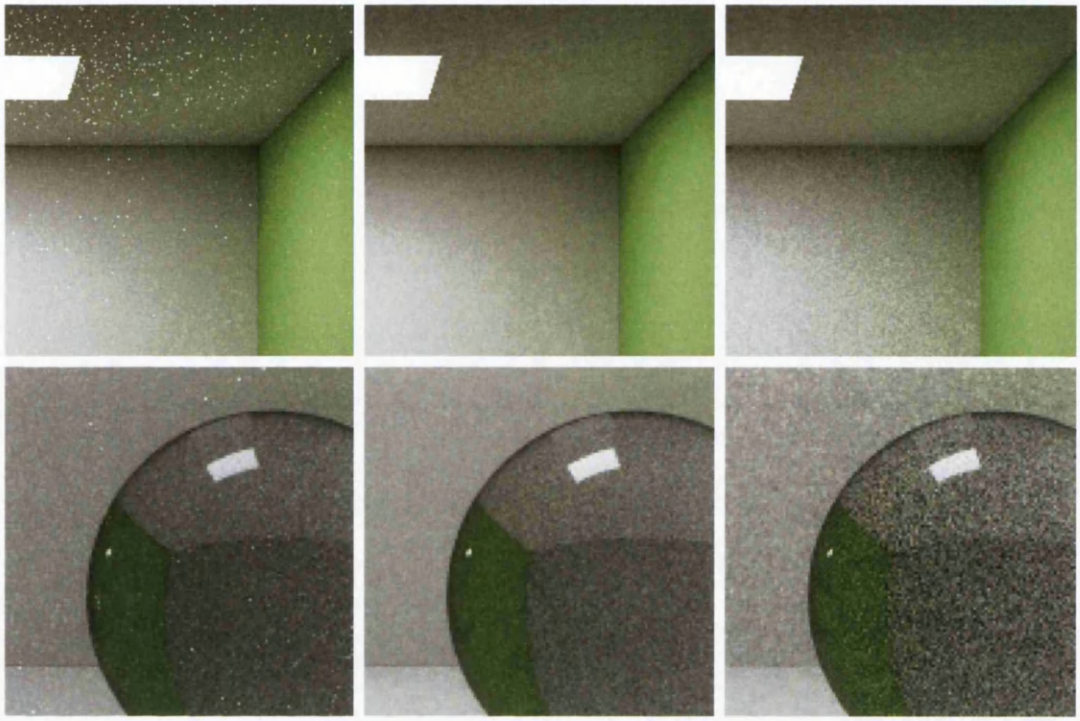


Figure 5.6: *Path tracing is affected by its poor evaluation of caustic paths. Our method does not introduce the lower frequency noise produced by SPPM visible around the image edges, or the higher frequency noise on the back wall and glass ball. Images for all three methods were rendered in 30 minutes*

5.7.2 Metallic Ring - difficult caustics

While the Cornell box provides a number of general lighting conditions and phenomena, this second scene aims to provide more challenging caustics for a more rigorous test of caustic lighting evaluation. This scene (Figure 5.8) contains a metallic ring sitting on a flat detailed surface within a walled environment. Illumination is provided by a small area light source.



Path Tracing Our Method SPPM

Figure 5.7: *Close-up images of the Cornell Box scene for each of the three techniques, cropped from Figure 5.6. Low photon densities increase noise levels for the SPPM algorithm, and variance is dependent on the surface orientation. PT struggles to cope with the sparse and high energy caustic lighting producing distracting speckles.*

Direct caustics are visible in the centre as a high energy cardioid caustic, and more subtle caustics outside the ring. These features are also reflected in the interior and exterior of the ring respectively, which is difficult for methods that do not rely on sample caching, such as unbiased MC methods. Due to the light falloff on the walls of the environment, these regions are relatively dark and produce soft diffuse gradients.

Path tracing cannot effectively render the cardioid caustic due to the small light source which is difficult to sample implicitly. Likewise, the reflection of such caustics also proves difficult, despite being an extension of the same paths the lack of coherence reduces the effectiveness of unbiased MC approaches.

SPPM has noticeable noise as a result of low photon densities on the wooden floor and brick walls. Our method evaluates all light paths effectively, producing diffuse lighting with similar quality to the path tracing image, and caustics comparable to those produced by SPPM.

5.7.3 Shapes - Reflective open environment

Finally, a scenario is presented that can be challenging for both camera path and light tracing methods (Figure 5.10). Due to its open nature, particle tracing techniques perform poorly,

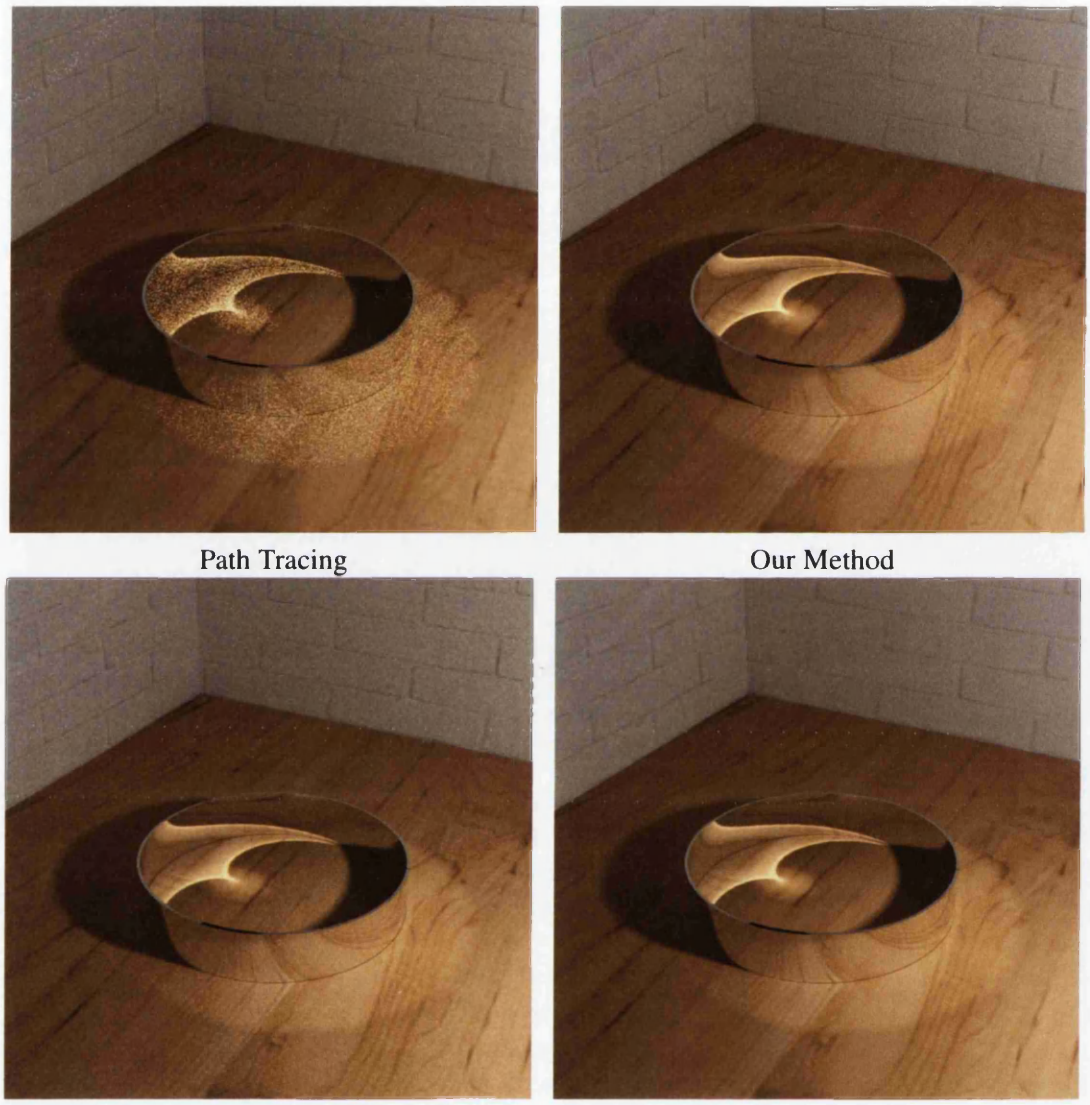


Figure 5.8: *Equal time comparison after 30 minutes between path tracing, the hybrid PTPPM method and SPPM. A low noise path tracing image is also provided for reference, which took many hours to render. Close ups of the same images show reductions in noise for our method compared to both methods*

allowing many emitted photons to miss the geometry entirely or be deposited in regions with low camera importance. An area light source illuminates a large plane containing diffuse, glass, and reflective objects. Caustics are visible directly as well as via reflections and transmission, thus posing a further challenge for unbiased MC techniques.

To highlight the improvements brought about during the early stages of rendering, under low sample counts, this scene has been rendered for four minutes as opposed to thirty. As pro-

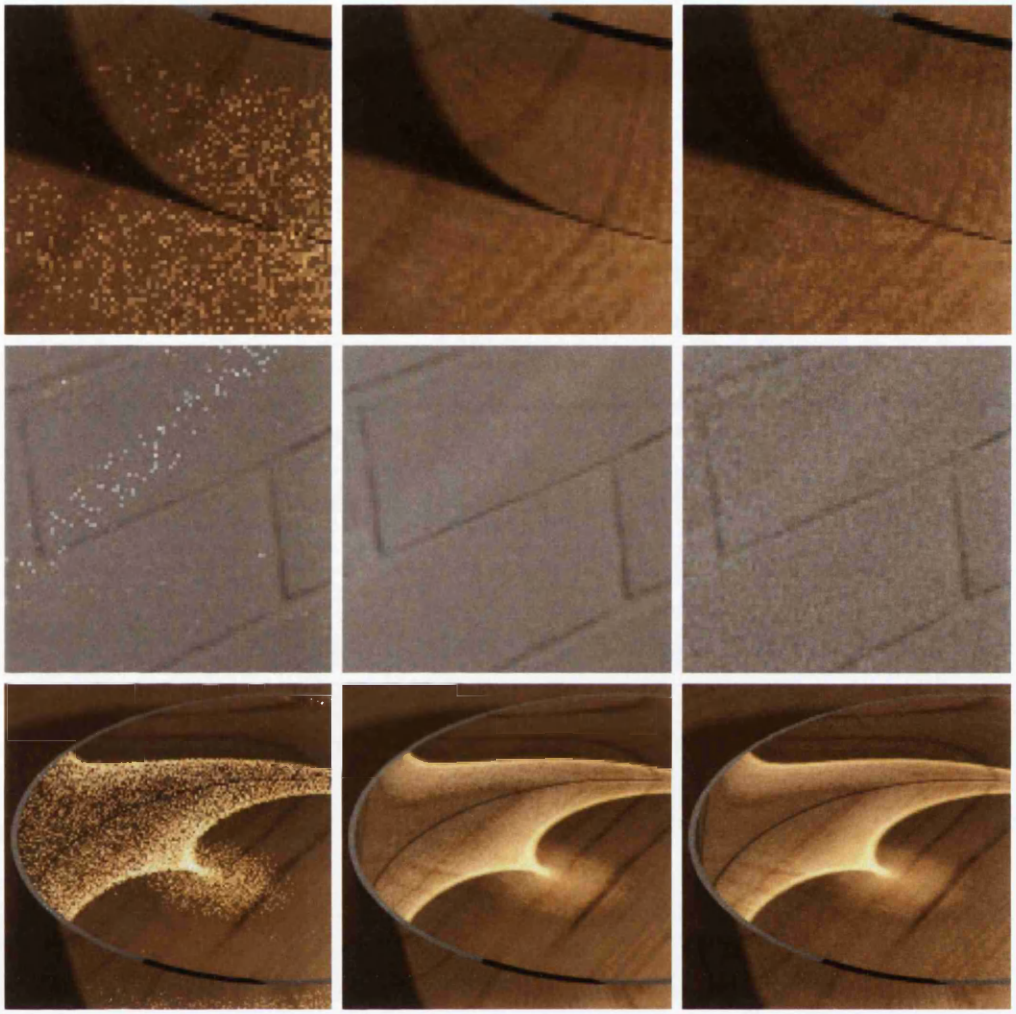
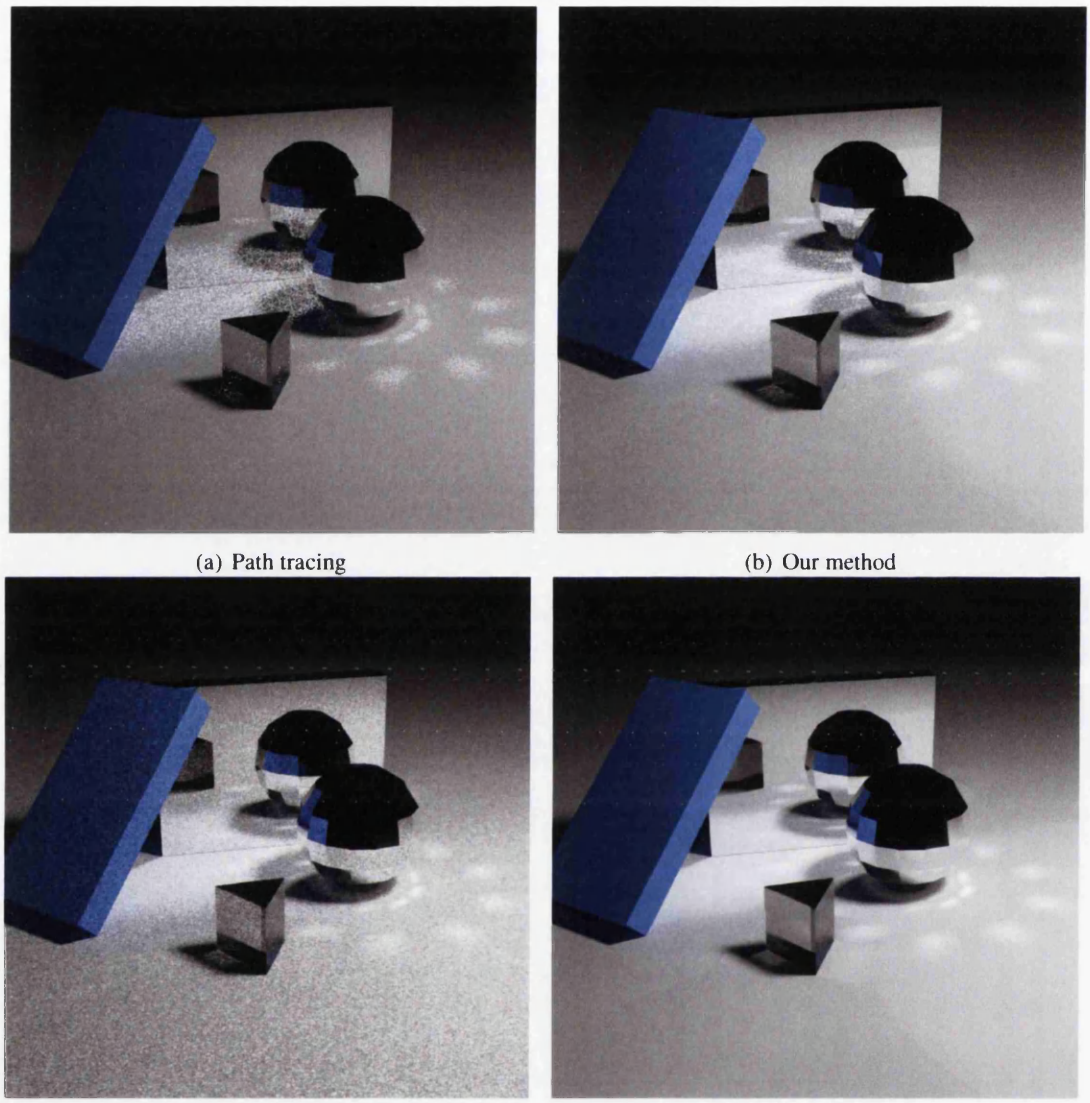


Figure 5.9: *Close-up images of the reflected illumination and shadowed regions from the Metallic Ring scene for each of the three techniques. Cropped from Figure 5.8.*

gressive techniques are often most useful for rendering previews, before generating production quality images, the early stages of rendering are where noise reduction can be most rewarding. This can be seen in Figures 5.14 and 5.13, where the error visible in the hybrid approach is markedly below that of previous techniques, and is localised around the more challenging areas of the image; the caustics.

Interestingly, rendering using path tracing provides lower overall error for this scene than SPPM due to the wide dispersion of photons traced from the light source, and the significant diffuse illumination across the image. Even with the similar photon distribution, the direct lighting at diffuse vertices provides vast improvements in the hybrid method.



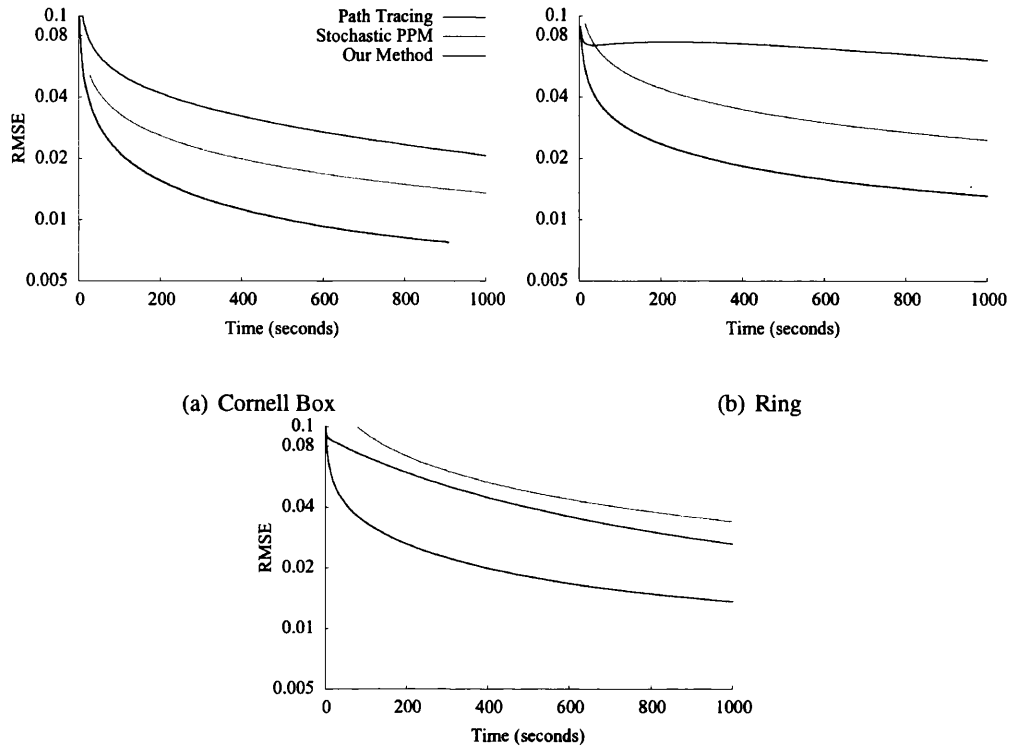
(a) Path tracing (b) Our method

(c) SPPM (d) Path traced reference

Figure 5.10: *Equal time comparison after 4 minutes of rendering between path tracing, our method and SPPM. This scene displays both direct and indirect lighting, in addition to a range of reflected and refracted light paths. Due to its open nature, both the indirect diffuse lighting on the blue box, and the caustic lighting seen via reflection and refraction are difficult to evaluate*

5.7.4 Quantitative comparisons

Computing RMSE values for the three scenes further demonstrates the effectiveness of the hybrid rendering technique (Figure 5.11). To preserve equality, these error calculations have been applied to the raw pixel data, before the application of tone mapping. In regions where diffuse



(a) Cornell Box

(b) Ring

(c) Scene

Figure 5.11: *Plots of* log *root mean squared error (RMSE) over time for the three scenes. The hybrid approach provides consistently lower error, and convergence is dependant on the relative complexity of the diffuse and caustic lighting.*

lighting presents more of a challenge, such as the Cornell Box scene and heavily shadowed regions, convergence of the hybrid approach follows that of path tracing. In scenes like the metallic rings, dominated by caustic lighting convergence resembles the rate of SPPM. In more mixed environments such as the final test scene, both caustic and diffuse lighting to converge at similar rates, with neither path sub space consistently dominating the rate of convergence.

Table 5.1 shows the number of caustic paths evaluated using the combined PT and SPPM technique in comparison to the standard MC path tracing renderer for the test scenes. Light tracing allows many more caustic paths to be evaluated which results in the vast improvements in rendering quality for such illumination. The final column shows the percentage of photons that are stored in the photon maps on average, compared to the same number of photons without the vertex contribution, path-space filtering. Thus, the size and memory requirements of the new photon mapping implementation are reduced to around 7%-15% in comparison to SPPM when used alone storing photons for the full global illumination solution. This dramatic decrease allows faster photon gathering as can be seen by the photon statistics in Table 5.2.

Path tracing dominates the runtime of the hybrid approach (PTPPM) for all scenes, due to the next event estimation performed at each camera path vertex to obtain radiance contribu-

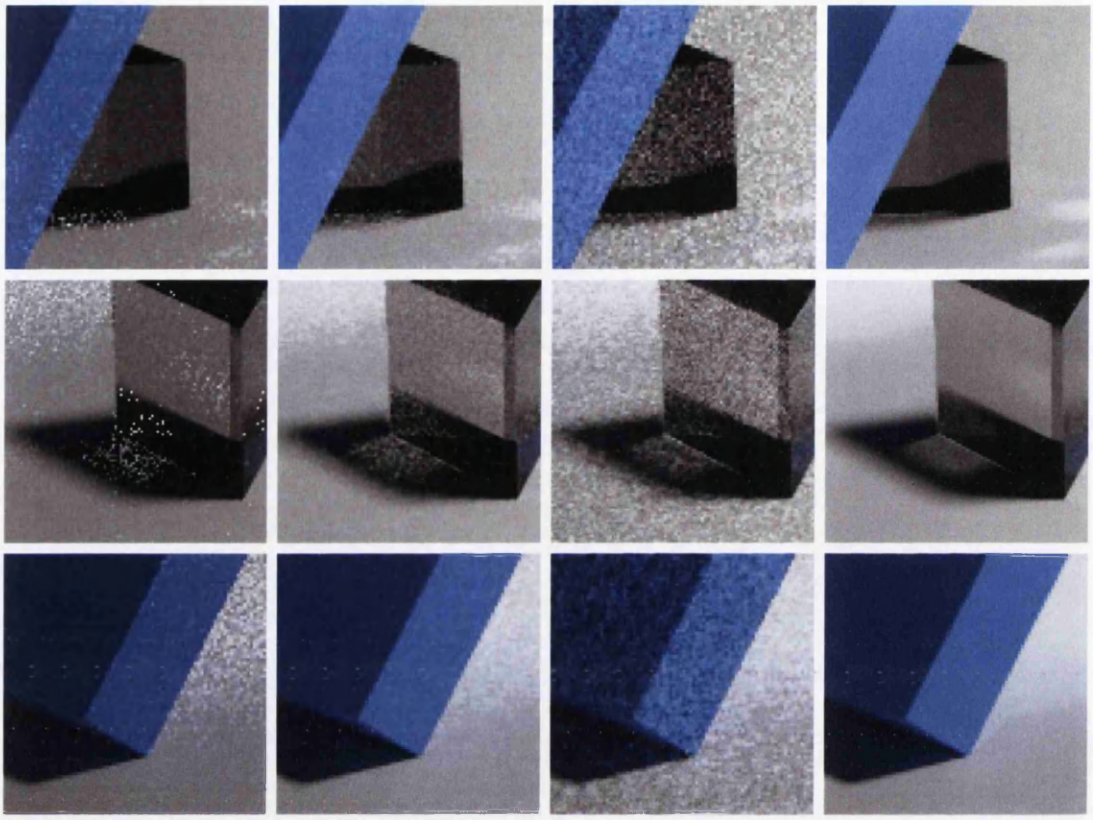


Figure 5.12: *Close-up images of the Shapes scene for each of the three techniques, cropped from Figure 5.10. Due to the open environment and difficult caustic lighting, neither PT or SPPM perform well under such conditions.*

tions, requiring many more rays than for a single photon path.

5.8 Limitations and future work

It can be seen that SPPM leaves significant dark portions of the image, localised over a small pixel region. Under sparse photon densities, insufficient photons exist in the footprint of the pixel for any given iteration to provide good estimates. Enlarging the kernel footprint may not be a feasible solution, especially where the local geometry is complex as additional photons may not be available. This establishes significant levels of variance across the image even during the initial passes. As the radius of the kernel shrinks, the problem is compounded since variance increases progressively in order to reduce the bias afforded by density estimation. Thus the absence of this initial bias results in high variance from the start, and the efficiency of the estimator is poor.

By allowing the use of unbiased point-to-point methods, the hybrid approach developed in

Scene		Iterations PT	Iterations PM	Path Tracing	Photon Tracing	Photon Gathering	Image Reconstruction
Cornell Box	*PTPPM*	1640	820	1420s	289s	70.5s	18.3s
(30 mins)	*PT*	2034	-	1800s	-	-	-
	SPPM	-	1282	271s	479s	1040s	12.4s
Metallic Ring	*PTPPM*	1172	586	1410s	325s	51.5s	12.4s
(30 mins)	*PT*	1446	-	1800s	-	-	-
	SPPM	-	929	370s	721s	689s	11.5s
Shapes	*PTPPM*	340	170	139s	56.3s	32.2s	5.05s
(4 mins)	*PT*	458	-	240s	-	-	-
	SPPM	-	186	63.5s	80.6s	92.0s	4.85s

Table 5.2: *Number of iterations performed by each algorithm, and the time spent in each stage, whilst producing the images shown in this section. Note that each iteration of PTPPM method evaluates two Monte Carlo paths, and gathers photons around one hitpoint. Path Tracing times for PTPPM include the path generation, regular expression filtering and diffuse radiance calculations. For SPPM it represents the time spent in the distributed ray tracing passes.*

this chapter can employ explicit sampling for important regions of the image without relying implicitly on the underlying photon density being sufficient. Conversely, in high energy caustic regions where the underlying density is high, photon mapping can provide accurate estimates using small filter kernels. This ensures estimators have low initial variance, and as the kernel bandwidth shrinks, the photon density is often sufficient to introduce variance slowly. Despite this, initial bias can still be a problem, as with the original (S)PPM technique, but ensuring that the kernel bandwidth is no larger than the ray footprint minimises the blurring of illumination across neighbouring pixels.

As with previous photon mapping approaches, performing density estimation on strong illumination boundaries can often over-estimate the radiance contribution to the pixel, due to proximity bias. High energy caustics can often contain sharp illumination features, or have strong borders with low density regions outside the envelope of the caustic. For progressive photon mapping, as the kernel decreases this proximity bias of the existing $N_i^c(S)$ photons is offset by those collected in subsequent iterations. However for pixels outside the boundary of sharp caustic features, as the radius of the kernel is reduced the number of additional photons $M_i^c(x_i)$ may be small in relation to $N_i^c(S)$. Relying on the ratio of these pixel statistics to dictate the radius reduction results in a slow decrease in bias, increasing its persistence during rendering. As the hybrid approach proposed here does not include the non-caustic photons, the difference in photon densities at such illumination boundaries is increased and thus the residual bias left behind by relying on this ratio of photon counts is also increased. By using the radius reduction technique of Knaus and Zwicker [KZ11], which is independent of the local statistics, reducing the radius by a fixed amount each iteration regardless of the number photons collected, the persistence of such bias can be reduced.

Omitting the diffuse photons reduces the overall photon density in the scene. Knaus and Zwicker [KZ11] show that the same convergence rate can be achieved by ignoring local statistics and reducing radius rates uniformly at each iteration, assuming a constant local photon density. However for low density regions, some iterations may not collect any photons, and the density is non-constant. Hence $M_i^c(x_i) = \alpha M_i^c(x_i) = 0$, and so the radius $R_i(S)$ will not be reduced if $N_i^c(S)$ is non-zero. Under these conditions, the progressive bias and variance trade off is affected since the radius is not reduced and the current flux estimate remains the same. Utilising the probabilistic method of Knaus and Zwicker will alleviate this, since both the radius reduction and flux correction are independent of the local statistics. Since the advent of the work described in this chapter, Kaplanyan and Dachsbacher [KD13] have discussed the problem of initial kernel bandwidth selection to reduce bias and variance. They develop a locally-adaptive technique to balance variance and bias to minimise error in a general PPM context, hence it is also applicable in this case. The core problem for density estimation still remains, but by intelligent choice of initial kernel bandwidth it can be minimised on a per pixel basis.

Given the percentage of photons stored in the photon map, performance for photon gathering is improved but the photon tracing step itself can become inefficient. This is largely due to the inclusion of indirect caustic sub-paths *LDS*, requiring an additional ray to be sampled and traced for each photon path before allowing termination (to identify *LDD* sub-paths). Such paths could be included in the path tracing stage for scenes with low intensity caustics as decided by the user.

Glossy materials are often problematic in a hybrid framework such as that presented here, and they can be difficult to sample with both unbiased MC methods and photon mapping approaches. The classification of glossy materials is non-trivial, as their representation and appearance can be arbitrarily close to either diffuse Lambertian surfaces or specular materials like polished metals. Dammertz et al.[DKL10] treat them as diffuse or specular based on a predefined user choice, affecting only the convergence of the algorithm not its visual appearance. Ideally, this would automatically identify the best technique to handle an individual path or sub-path based on the properties of the path vertex in order to reduce error. The addition of detail such as specular maps can make an absolute choice impractical and more complex BRDFs are highly dependent on the incident angle of the rays.

Following the publication of the contributions described in this chapter [DJM12], Georgiev et al.[GKDS12] and Hachisuka et al.[HPJ12] have independently presented similar techniques for combining BDPT and SPPM using multiple importance sampling to automatically weight the two techniques at each path vertex, reducing variance. Difficulties in such an approach arise due to the different path spaces that similar contributions occupy, when using point to point connections and density estimation. To connect a camera and light path of lengths i and j respectively, vertex connections can be made explicitly for BDPT resulting in a total path of length $i+j$. For density estimation to construct a similar path, the vertices of both paths need to be neighbouring in Euclidean space, according to the gather kernel radius. Thus a path of length $i+j+1$ is required, and in order to combine such techniques on a per-vertex basis requires a common space from which to compute path densities. Extending the path space of the BDPT approach can be achieved by randomly generating a vertex within the kernel radius of the density estimation technique, and including it in the density of generating that

path contribution. Alternatively, the space of density estimation techniques can be reduced by merging neighbouring vertices [GKDS12], matching the space of MC path estimators.

A weighted approach improves the evaluation of glossy materials and reduced visible spiked noise in images, which are robust under a number of difficult scenarios. This comes at the cost of high implementation complexity, not only subsuming BDPT and SPPM, but requiring additional care to ensure the MIS weights for each technique are computed correctly, and presents further difficulties for efficient implementation.

The complexity of such weighted approaches makes the use of Markov-chain MC techniques difficult, and such a solution is yet to be established. By combining MC and photon mapping techniques in a simplified but effective hybrid framework, the PTPPM technique presented here can easily utilise Metropolis-light transport for difficult diffuse illumination, and equivalent techniques for photon mapping [HJ11] can improve caustic illumination.

The algorithm presented here represents a concept that is equally applicable to any Monte Carlo orientated method such as bidirectional path tracing or Metropolis light transport. Utilising BDPT in place of uni-directional path tracing could provide significant performance improvements and reduce the overhead of a hybrid technique, since photon paths that do not contribute to the photon map can be utilised for the diffuse lighting computation. Similarly, caustic paths generated from the camera whose contributions are skipped could also be reweighted and added to the photon map although in practice the number of additional photons would be relatively small for anything but simple cases. Utilising BDPT for direct caustics could also improve convergence due to the light tracing sub paths, however this would risk re-introducing the high energy spiked noise that is being eliminated.

A desirable addition would be to allow automatic adjustment to the number of Monte Carlo generated and photon traced paths during rendering, improving convergence for scenes with difficult caustic or diffuse lighting. Building upon the bias and error estimation work of [HJJ10] may allow a more rigorous estimation of error present in each technique, allowing an automatic adjustment. Care would need to be taken however to ensure that portions of path space were not under sampled prematurely, if they have not been explored sufficiently and their variance is not representative of the actual integral.

From an implementation perspective, the current efficiency of the estimator has room for significant improvement. Subsuming both path tracing and photon mapping progressively results in a wealth of information being available to guide subsequent passes without the need for pre-processing. Data from each progressive pass can be used to guide the evaluation of the opposing integral. Previous techniques such as significance caching [BRDC12] and photon driven importance sampling [Jen95] could provide useful information to both passes improving the illumination estimates. Techniques for more efficient photon mapping can also be applied, such as the use of projection maps [Jen01] to refine direct caustics and a technique applicable to the indirect distribution of photons could also be advantageous, may require preprocessing of the geometry and the inclusion of visibility queries. A technique similar to shadow photons [Jen01], in which a sparse set of direct lighting photons (which are already computed and discarded in the implementation described here) are stored and queried could improve efficiency by allowing probabilistic evaluation of direct lighting computations at each vertex.

The convergence of hybrid methods such as that presented here and the similar works by Georgiev [GKDS12] and Hachisuka et al.[HPJ12] are still limited by the lower convergence

rate of SPPM, for areas of the image where SPPM is used for sample evaluation. As the radius is reduced, evaluation by path tracing techniques can be favoured and thus in the limit the solution will become unbiased. Optimal combinations of such a complex framework is yet to be established, and provides an interesting foundation for future work. The use of path vertex data as discussed in this chapter, and the concept of path classification can still benefit such techniques to improve convergence by adjusting the balance between light and camera paths to improve specific types of illumination.

Finally, combining unbiased Monte Carlo and density estimation approaches has been shown to benefit the rendering of global illumination, but still inherits problems from the original contributing techniques. Bias reduction for density estimation is still problematic, and errors due to poor distributions over complex geometry and the results of proximity bias still impact the resulting images. High frequency noise and effective techniques for evaluating the radiance scattered from glossy materials are still problematic. Recent work by Jakob and Marschner [JM12] to enable more effective Markov-chain mutations through complex specular and near-specular light transport improving rendering through the exploration of path-space manifolds.

5.9 Conclusions

Achieving high efficiency whilst maintaining robustness is a desirable but difficult to attain property for computer graphics algorithms. It has been shown in the past that a single specific algorithm can often solve a particular light transport problem more efficiently than a generalised one [DKL10, KK04, DWB+06]. A novel multi-pass progressive algorithm has been presented that combines the benefits of both Monte Carlo path based and progressive photon tracing methods via path space filtering. Though separable path space filtering methods have been used before for full global illumination [DKL10, BAJ08], the PTPPM algorithm has the advantage of being both progressive and accurate, allowing convergence to the correct solution or until the desirable level of quality is achieved.

The core of this chapter has been to introduce a hybrid density estimation and pure Monte Carlo path tracing techniques in a progressive framework which, via quantitative RMSE analysis, has shown to be more effective than the individual techniques it subsumes. The resulting work incorporates a number of contributions:

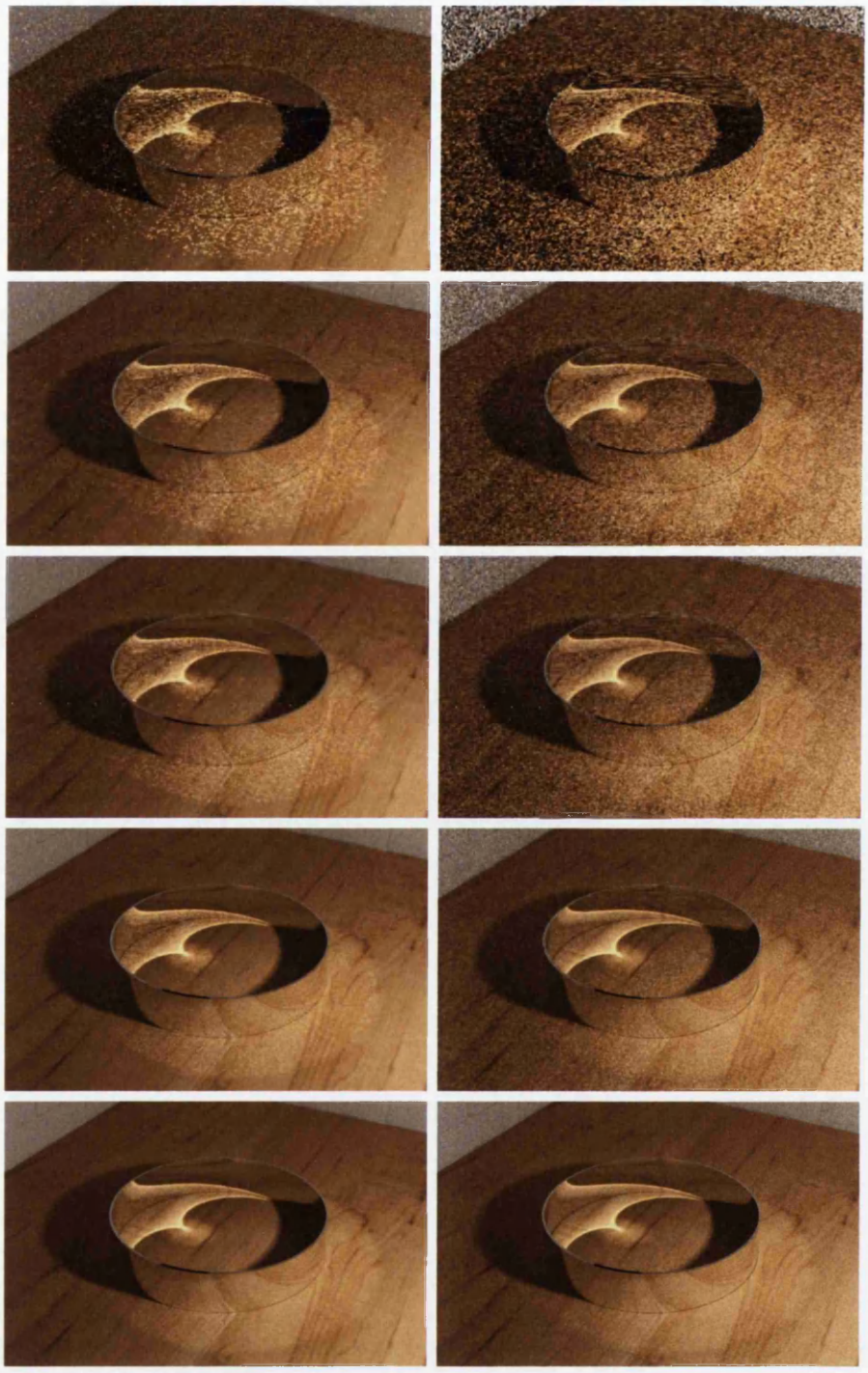
- By rearranging the rendering pipeline, radiance accumulated at path vertices can be classified based on their surface interactions, and evaluated lazily to improve the performance of path tracing techniques.
- Support for visually significant *SDS* paths has been combined with unbiased rendering of general illumination to reduce overall variance and improve accuracy.
- High efficiency is achieved by significantly reducing the memory and computational costs of density estimation techniques for caustic lighting.

Using path-vertex data has allowed the identification and division of caustic illumination that causes high variance in global illumination path tracing and insight into the characteristics

and distribution of such path space. For a given estimator, it can be seen that some regions of path space are highly problematic whilst others are relatively noise free. Figure 5.3 illustrated this in the context of path tracing, and motivated the work in this chapter.

Aside from quality and speed, the two driving forces behind computer graphics, the ease of use and toolkit available to non-specialist end users such as artists, designers, and engineers must be addressed in order to push global illumination into the widespread use it deserves. As such, investigating hybrid approaches like that which has been presented here to improve convergence is an important step towards developing a framework that can cope with all manner of illumination effects equally well. This reduces the knowledge required by the end user, eliminating the need for questions such as "Which algorithm do I choose for this particular scene?" or "What does this parameter actually do?". Hence, eliminating the experimental tweaking of parameters in order to achieve good results allows users to focus instead on their primary goal, be it creative design or scientific analysis. Hybrid approaches such as the work presented in this chapter go some way to making accurate global illumination a universal tool, instead of an expertise.

This work focussed on utilising path data in high intensity caustic lighting and complex specular interactions that are handled poorly, if at all, by unbiased MC methods. However, in a more general case such path vertex data and its analysis may provide further benefit to achieve noise free rendering. Even within a single mode of light transport, such as non-caustic illumination, path vertex data can be used to isolate important features in the image.



PTPPM SPPM

Figure 5.13: *Images showing the progressive rendering methods at equal times. PTPPM results in visually more pleasing images compared to stochastic progressive photon mapping, especially at lower sample counts. From top to bottom approximately 4, 8, 16, 64 and 256 seconds respectively.*

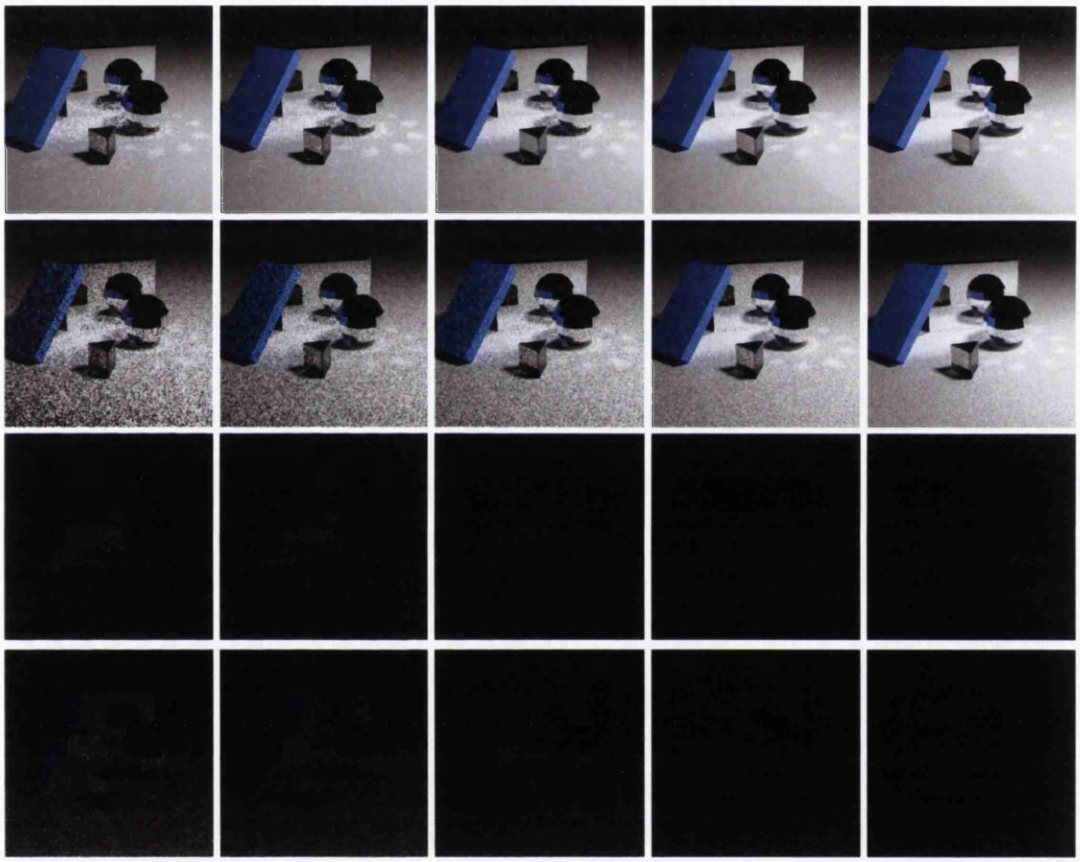


Figure 5.14: *Images showing the progressive rendering methods at equal times the novel PTPPM method (first row) produces in lower overall error than SPPM (second row). From left to right the images were rendered in approximately 4, 8, 16, 64 and 256 seconds. The third and fourth rows display the pixel-wise error for PTPPM and SPPM respectively, compared to the path tracing reference image in 5.10*

Chapter 6

Irradiance-Aware filtering for Monte Carlo rendering

Contents

6.1 Introduction **119**
6.1.1 Irradiance awareness in image based filtering 120
6.1.2 Contributions 122
6.1.3 Poisson Distribution 123
6.2 Rendering and sample clustering **124**
6.2.1 Cluster formation 127
6.3 Noise removal and filtering **132**
6.3.1 High intensity sample rejection 133
6.3.2 Illumination preserving filter 139
6.3.3 Pixel radiance computation 146
6.4 Results and discussion **149**
6.5 Limitations and future work **156**
6.6 Conclusions **158**

6.1 Introduction

Following the work proposed in the previous chapter, the concept of utilising vertex data to enable variance reduction has been established. Despite the improvements in visible noise and quantitative error made evident in the previous chapter, convergence for high quality rendering techniques is still inherently slow and produces visible noise early on. By visualising vertex contributions as a series of distinct layers based on the number of surface interactions (Figure 6.1) a number of observations can be made about the distribution of noise and characteristic illumination features in the image.

In environments with largely non-specular BRDFs, common in the real world and where the nature of scattering is non-deterministic, the dimensionality (and hence the complexity)

of the integral increases with vertex depth. Intuitively, the number of possible directions from which illumination could arrive increases, hence each additional path vertex expands the search space and thus more samples are required to provide similar levels of variance. This can be explained by looking at the formulation of path densities during path construction. The extension of a path from a point x to x' is dependent on the PDF $p(x')$ used to select an outgoing direction ω' over the hemisphere at x, typically via PDF importance sampling based on the BRDF and local geometry. As a path $\bar{x}_k = x_0, ..., x_k$ is extended from x_0, the probability of sampling the next vertex is dependent on the probability of all previous vertices, thus the path density up to x_k is given by:

$$p(\bar{x}_k) = p(x_0)\, p(x_1) \ldots p(x_k) = \prod_{i=1}^{k} p(x_i)$$

where $p(x_0)$ is the PDF associated with sampling the path origin (ie: the pixel location on the image plane, or point on an emitter). To be valid, it must hold that $p(x) < 1$ for any x and ω', unless p is deterministic (a purely specular interaction) where $p(x) = 1$ may hold. Thus for non-specular interactions, as k increases $p(x_k)$ is dependent on exactly $k-1$ PDFs where $p(x) < 1$ for all x_0 to x_k. None of these PDFs can fully account for the incident radiance (irradiance) at x_k, thus in general the potential variance of the estimator increases with higher dimensions [1]. As a result, longer paths increase the chance of background noise being introduced into the image. Additionally, the variation of each of the $k-1$ choices for $p(x) \in (0,1]^{\mathbb{R}}$ at each vertex adds to the total variation in path density $p(\bar{x}_k) \in (0,1]^{\mathbb{R}}$ for similar paths of equivalent luminance.

Another observation from Figure 6.1 is that the frequency of illumination features changes with path depth. Due to the dispersive nature of wide BRDFs, the presence of strong features in the illumination are reduced as energy is spread more uniformly around the scene. As a result, phenomena become softer displaying lower frequency details at longer path lengths. Thus the noise reduction power of techniques operating over composite path radiance from all path space (as opposed to per-vertex contributions), are restricted by all illumination features across the composite path space.

This also applies to features in visible texture maps which form an essential part of the final radiance measurements. Noise in the incident illumination is present irrespective of the texture detail on the surface, and thus filtering the final sample radiance inclusive of such detail provides additional restraints, potentially reducing the capability of the filtering scheme. Isolating both the signal and underlying noise of such features individually as a further tool for noise reduction is the motivation for the novel contributions throughout the following chapter, culminating in a noise reduction framework for progressive rendering.

6.1.1 Irradiance awareness in image based filtering

In order to make use of these observations, a vertex level approach is explored for image-based filtering; a wide and varied field of noise reduction techniques. This section will provide

[1] Note that only if the incident illumination were already known could a PDF introduce no additional variance to the estimate, in which case sampling is unnecessary anyway since the solution must also be known.

(a) Primary diffuse lighting (b) Single diffuse interaction (c) Multiple diffuse interactions

Figure 6.1: *An example (taken from Figure 5.3) showing distinct sets of path tracing contributions, identified based on vertex interactions and path length. From this, features of the illumination can be identified and attributed to dimensions of the integrand, which is used to improve noise reduction techniques.*

motivation for the application of this work to image-based filtering, and the principles behind such techniques for noise reduction in Monte Carlo path tracing.

To accurately estimate the measured radiance through a pixel in a path based Monte Carlo renderer, the radiance from all surface points visible through that pixel must be evaluated:

$$P_{(i,j)} = \int_S \int_\Omega L_e(x,\omega) + L_r(x,\omega) \mathrm{d}x \mathrm{d}\omega, \tag{6.1}$$

for the continuous set of all visible locations in the scene S and the set of directions towards the camera Ω from S. To estimate L_r, further integration at x is required to account for the radiance arriving from all directions ω' over the hemisphere:

$$L_r(x,\omega) = \int_\Omega L_i(x,\omega') f_r(x,\omega,\omega')(\mathbf{n}\cdot\omega')\mathrm{d}\omega' \tag{6.2}$$

where L_i is the incident radiance arriving at the surface from ω'.

Typically, these integrals are slowly varying, that is for small changes in x and ω, the change in $L_r(x,\omega)$ is also small. Such similarity exists across neighbouring pixels in the image plane, since pixels are a discrete reconstruction of the continuous 2D image plane, commonly translating into a small spatial distances (in x and ω) between ray hit points in world space. The incident illumination L_i at each hit point thus changes slowly in diffuse environments with simple geometries. Filtering techniques rely on this pixel coherence, exploiting the similarities in the signal of neighbouring integrals to reduce the incoherent noise that varies across them, as a result of stochastic sampling. This approach can provide effective noise removal in regions where neighbouring pixels have well correlated integrals.

For non-planar geometry and at geometric edges, hit points from neighbouring pixels exhibit large or sudden changes in the local geometry, resulting in significant differences in their illumination integrals. Additionally, the coherence of the scattered radiance L_r between nearby

hit points is more sensitive under narrow BRDFs, which preserve small changes in the incident illumination. The presence of texture detail can also reduce pixel similarity and degrade coherence quickly across the image plane due to fine detail on visible surfaces.

For such regions which lack coherence, low variance sources such as geometry and texture buffers can be employed to help preserve visually distinctive edges in conjunction with filters such as the cross-bilateral (recall the discussion in Section 4.4). Image based techniques have been popular due to their fast performance but often sacrifice visual quality. The resulting algorithms can produce filtering artefacts, loss of high frequency features, or are limited to certain forms of light transport; wherein lie the strengths of Monte Carlo methods. However relying on low variance sources cannot be applied to identifying pixel correlation with respect to illumination features, since they are the source of high variance that is the target of the filter. As a result, a number of existing techniques do not filter the direct lighting or handle high frequency environments [BEM11, SKBF12], where the introduction of bias has a significant impact on the resulting image, blurring shadow edges and reflections. However, such illumination still brings noise to the image that can be distracting especially in shadow penumbra, so filtering is still desirable.

6.1.2 Contributions

In this work, a novel framework for the dynamic storage and filtering of Monte Carlo samples is presented that effectively removes noise whilst preserving high frequency features in the incident illumination and texture detail. Utilising path length and vertex interactions, sets of per-pixel clusters are maintained, representing the radiance and incident radiance integrals of each pixel. This results in a layered clustering framework that represents the pixel integral.

Treating path tracing as a Poisson process,and clustering in this way allows the frequency of sample occurrences in a pixel to be compared with the occurrence of similar luminance samples in its neighbourhood. From this, two advantages are obtained. First, high energy noise in the incident radiance on visible surfaces can be identified and temporarily removed, which is otherwise difficult to handle without strong artefacts. Second, clusters of similar luminance can be compared across an image-space filter kernel, allowing weights to be derived for convolution filtering, respecting similarities in the incident radiance. Combined with the geometric edge detection provided by depth and normal buffers used by more established cross-bilateral filtering techniques, bias is reduced across edges from both high and low variance sources.

Furthermore, this can be done so as to preserve complex texture detail since the incoming radiance can be filtered as opposed to the final radiance arriving at the camera. Though not in itself novel, combined with the ability to decompose illumination based on high variance characteristics, the data available within a given filter kernel and the correlation of illumination integrals can be maximised across regions with multiple complex features. The final pixel radiance is reconstructed from the filtered irradiance, statistical BRDF estimates, and texture data all of which are obtained cheaply during the rendering pass.

The method proposed in this chapter effectively filters noise from the incident illumination, whilst preserving hard gradients and edges that are otherwise difficult to identify, pro-

viding important visual cues and realism for complex materials and light transport. The main contributions of this chapter can be summarised as:

- Introduction of a novel layer-based clustering framework, that compactly and accurately represents changes in the incident illumination (irradiance).
- Image based irradiance filtering, enabling the preservation of texture-based and illumination-based features without the need to store BRDFs or local geometry data.
- A probabilistic, discontinuity aware image-space filtering algorithm that is sensitive to illumination from multiple overlapping sources.
- The introduction of the Poisson probability distribution to noise removal, producing a statistical framework useful at low sample counts.
- Improvements to high intensity noise removal techniques, including the ability to tune results post-render, and suitable as a pre-process for image-based filtering.

The work presented in this chapter has undergone peer review and has been published and presented at the *Computer Graphics International 2013* conference and was further selected to appear in a special issue of *The Visual Computer* journal [DJ13].

6.1.3 Poisson Distribution

Before delving into the details of irradiance aware filtering, a brief introduction is provided to the Poisson probability distribution. This forms an important part of the procedure for identifying statistical differences between the illumination integrals of neighbouring pixels, so that they can be preserved.

The Poisson distribution is a discrete probability distribution expressing the likelihood that a given number of events will occur within a specified interval. It is based on the assumption that such events occur with a known average rate and are generated independently of the interval since the last event. Given an average of λ events have been observed in a fixed interval and that the process or processes generating such events are random with respect to the frequency of occurrence, the probability of observing exactly k events is given by the probability mass function:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{6.3}$$

where e is Euler's number, the base of the natural logarithm. Examples of the Poisson distribution for various mean occurrences are shown in Figure 6.2. Notice that the Poisson distribution is a skewed distribution, accounting for the absence of probabilities for negative k. As λ moves away from zero, the Poisson distribution becomes symmetric, following the normal distribution with mean λ and standard deviation $\sqrt{\lambda}$:

$$\frac{1}{\sqrt{\lambda}\sqrt{2\pi}} e^{-\frac{(x-\lambda)^2}{2\lambda^2}}$$

The poisson distribution can be used to model and predict a number of processes that are based on random events, ranging from prediction of defects in manufacturing, or the number

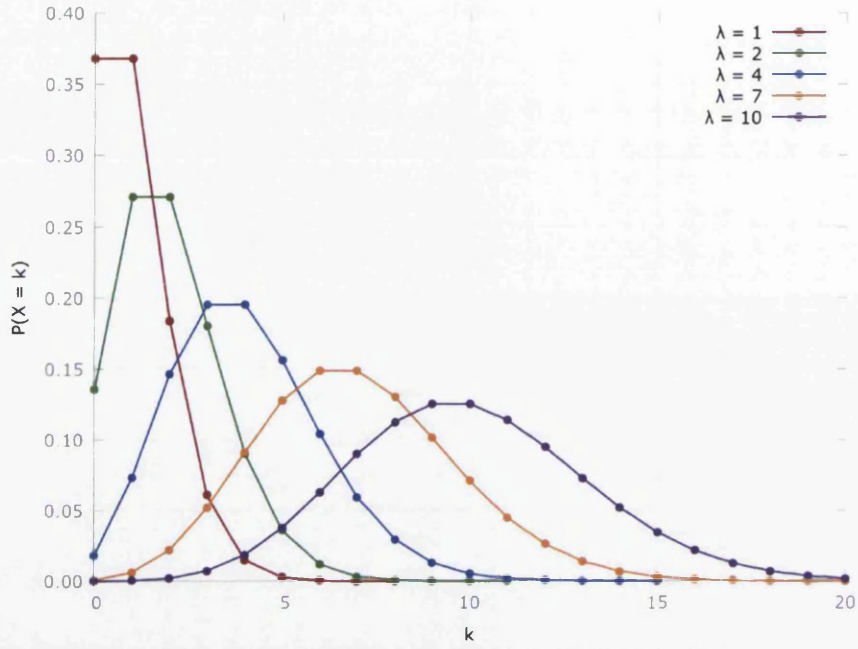


Figure 6.2: *Plot of the Poisson distribution, showing the probability mass function $P(X = k)$ for various mean number of observed events, λ. Notice as λ increases the distribution resembles the binomial and standard normal distributions. The lines on the graph do not represent the distribution but are a visual guide as to its behaviour.*

of calls arriving at a call centre in a given time period. A famous and classic example is Bortkiewicz's 1898 study of men kicked to death by horses in the Prussian cavalry, which was found to exactly follow the Poisson distribution. Further details and derivations of probability distributions including the Poisson approximation can be found in many texts on mathematical probability, such as [Gor97].

The Poisson distribution is especially useful when modelling rare events, where a large number of events may be observed as a result of independent processes, but the generation of each event itself is rare, such is the case in Monte Carlo methods. Samples are generated stochastically during rendering, and despite making use of stratification and importance sampling techniques, the generation of a given sample is not influenced by previous samples.

6.2 Rendering and sample clustering

In order to describe the proposed filtering technique, the rendering procedure is first described, with focus on obtaining the vertex contributions. The rendering is divided into the path generation and radiance evaluation phases, as in the techniques presented in the previous chapter (recall Section 5.5), allowing for each vertex contribution to be identified. A path $\bar{x}_k$ of length k generated using MC sampling therefore returns a radiance contribution for each vertex v for

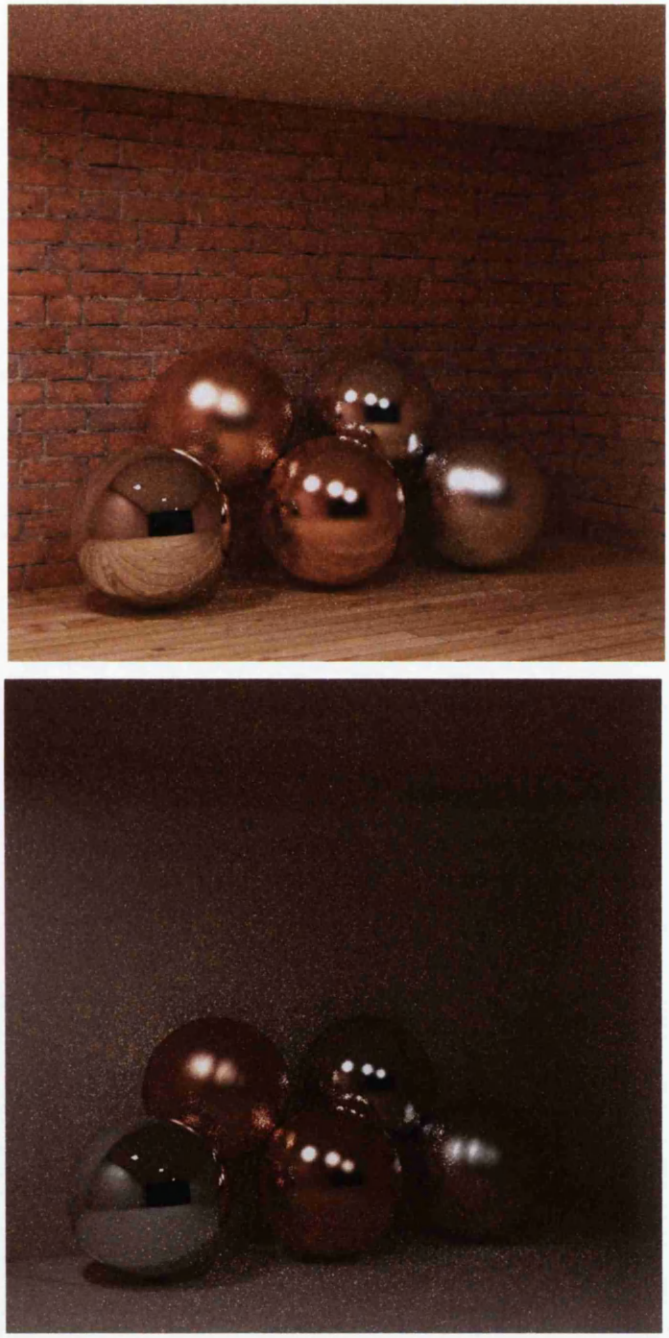


Figure 6.3: *Path tracing contributions returned by the renderer consist of the unbiased radiance estimates (top) as produced by a standard path tracer, and the incident radiance arriving at the first diffuse vertex (bottom). Notice particularly the absence of texture detail in the sharp reflections.*

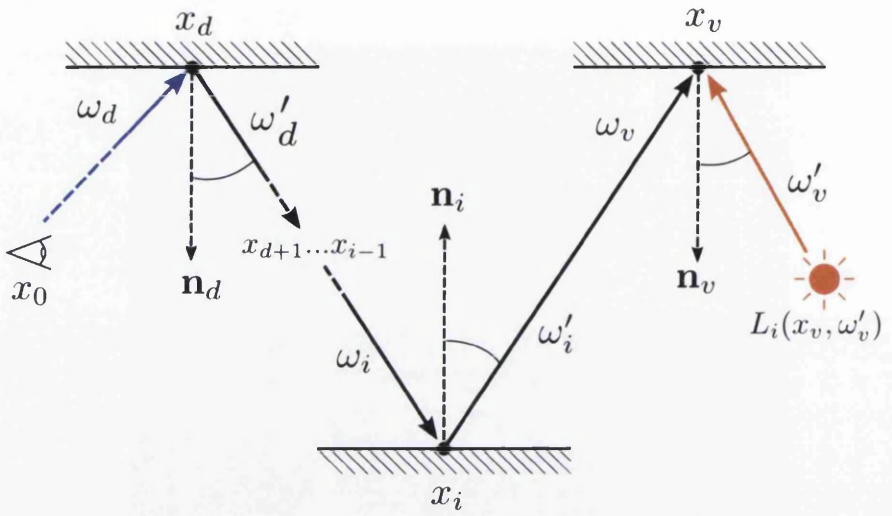


Figure 6.4: *Geometry illustration for the radiance and irradiance calculations in Equations 6.4 and 6.5. Radiance arriving at a vertex v is scaled by the path throughput $T(x_d \rightarrow x_v)$ from the first diffuse vertex x_d, omitting the BRDF and texture reflectance at x_d.*

$0 \leq v \leq k$ corresponding to the path space $\Omega_{\bar{k}} = \Omega_1, \Omega_2, ..., \Omega_k$ respectively.

In this work, the previous formulation is adapted such that each vertex contribution is further broken down to make use of readily available intermediate data. From each vertex contribution v, a pair of values is returned by the path tracer, representing the radiance and the incident radiance arriving at the first non-specular vertex x_d (Figure 6.3). The radiance contribution for a vertex v can be defined as:

$$R_v = L_i(x_v, \omega'_v) \prod_{i=1}^{v} f_r(x_i, \omega_i, \omega'_i)(\mathbf{n}_i \cdot \omega'_i) \tag{6.4}$$

where $L_i(x_v, \omega'_v)$ is the radiance arriving at x_v directly from the light source in the direction ω'_v (Figure 6.4). Since this is a product of coefficients, removing some terms from Equation 6.4 allows partial estimates to be computed, omitting certain dimensions of the rendering integral to obtain useful data independent of the remaining coefficients. In other words by removing the influence of certain parts of the integral which change rapidly, such as direct illumination or surface detail for example, similarities in the remaining radiance information can be shared. The incident radiance I_v for a vertex contribution v is obtained during the calculations required to evaluate the radiance R_v. This is somewhat similar to the radiance carried by photons in the photon mapping algorithms discussed previously, ignoring the effects of the BRDF and texture detail at the first diffuse vertex of the path. Thus the incident radiance as it arrives at x_d can be obtained by omitting $f_r(x_d, \omega_d, \omega'_d)$:

$$I_v = (\mathbf{n}_d \cdot \omega'_d) \\ L_i(x_v, \omega'_v) \prod_{i=d+1}^{v} f_r(x_i, \omega'_i, \omega_i)(\mathbf{n}_i \cdot \omega') \tag{6.5}$$

where d is dictated by preceding specular vertices i for $0 < i < d$. The premise for ignoring initial vertices before d is that specular interactions, due to their Dirac distributions, are more deterministic in nature and add little variance to the integral directly. The goal of the radiance separation is to isolate only the incident illumination; where the main source of variance lies.

The relationship between R_v and I_v for a given vertex contribution can be formulated as:

$$\frac{R_v}{I_v} = f_r(x_d, \omega_d, \omega_d') \prod_{i=1}^{d-1} f_r(x_i, \omega_i, \omega_i')(\mathbf{n}_i \cdot \omega_i') \tag{6.6}$$

which will be revisited later in more detail when discussing the filtering procedure in Section 6.3.3.

The result of this radiance separation is that high frequency detail resulting from the use of texture maps is removed. This detail varies quickly between pixels, regardless of the incident illumination, and so removing it from the radiance evaluation allows it to remain unaffected by any filtering performed over I_v; the primary source of the high variance.

As discussed in the introduction of this chapter, features in the incident illumination can also be isolated based on the dimension of the integral they lie in (for example, the path length). Bearing this in mind, the vertex contributions for the first three vertices following x_d in the camera path are maintained as separate contributions, forming the three layers of the filtering framework (Figure 6.5). Ignoring these vertices with respect to path length ensures that direct lighting arriving on non-specular vertices, of the form ES^*DL, is stored within the same layer of the clustering framework, since it evaluates the same integral over the diffuse vertex.

Although complicating the clustering process this improves sample separation and allows for easier detection of changes in illumination during filtering. By separating path contributions, each layer can be filtered independently, regardless of the illumination features and noise present in the remaining layers.

6.2.1 Cluster formation

Each path generated and evaluated during the rendering step produces a pair of radiance measurements for each of the three layers, as depicted by Figure 6.5. In order to make use of such contributions across the image plane, and to do so in an efficient manner a clustering scheme is developed. The design of this clustering framework is such that it can:

- Isolate visual discontinuities in the pixel integrals,
- Allow the efficient use of image-based convolution filters over the incident radiance,
- Maintain the unbiased radiance and its relationship with the incident radiance contributions (R_v and I_v respectively),
- Operate alongside a progressive renderer (ie: handle an arbitrary number of samples),
- Minimise the requirements for storage and computation during rendering.

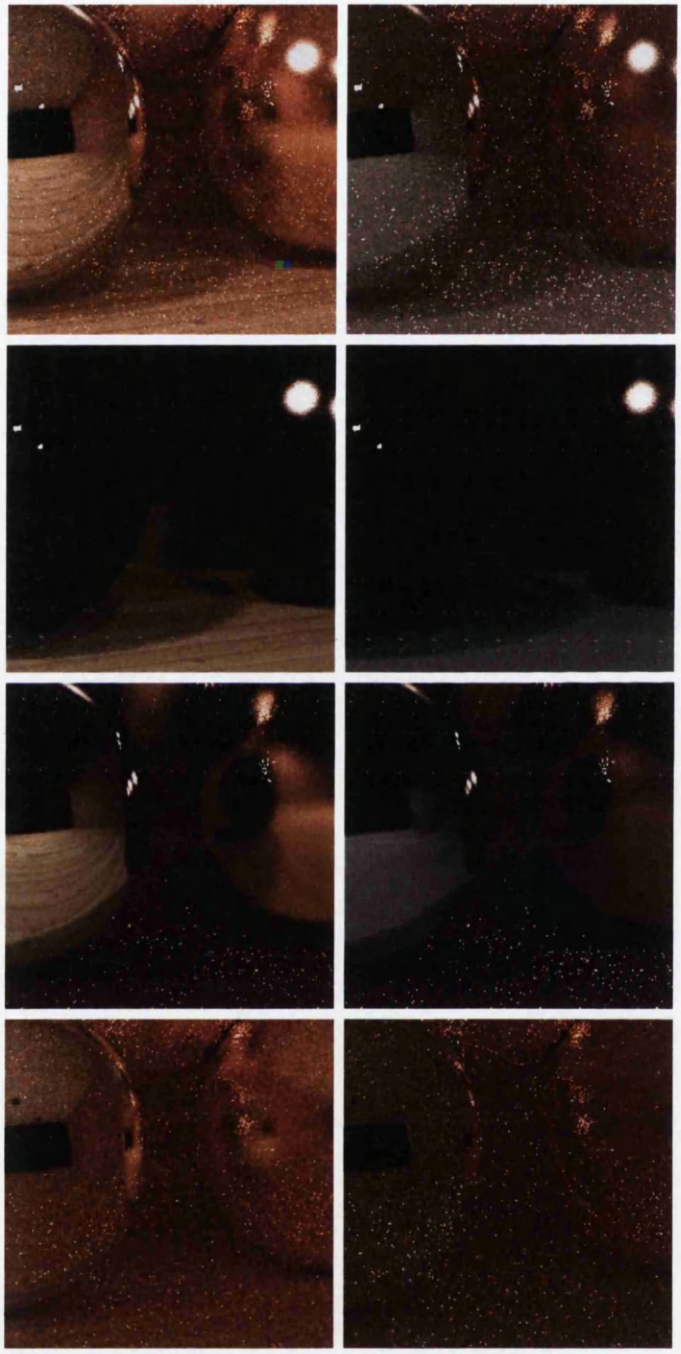


Figure 6.5: *Example of the layers in the clustering framework, enabling high frequency texture detail and changes in the illumination to be isolated simultaneously, based on the features at each path vertex.*

In this section, the resulting clustering approach is presented that tries to achieve these aims, building on the layer based radiance contributions derived in the previous section.

Clustering is performed on a per-pixel basis, producing a set of clusters that each represent a contiguous range of the pixel integral, whilst remaining disjoint from one another. Thus, each contribution is included in exactly one cluster. A "cluster set" or "set of clusters", denoted $C_0,...,C_{n-1}$ refers to the n clusters that represent all the samples evaluated so far (from the current and all previous iterations) for a given pixel at a given layer. For example, the three red clusters under $v = 1$ in Figure 6.6 are a set of clusters, C_0, C_1, C_2, where $n = 3$.

Since the goal is to isolate discontinuities in the incident illumination (where variance is highest) clustering is performed using incident radiance; the I_v measurements for each vertex contribution. Geometric and textural changes are not of interest at this point, at they can be resolved by the use of low variance range buffers, similar to those used in the image based filtering techniques previously discussed in Section 4.4.5.

The human visual system is generally more sensitive to changes in luminance than in chromaticity [Gla95]. As a result, preservation of the visual discontinuities in the clustering model is achieved using the 1-dimensional space of luminance, via the conversion of I_v to the perceptual CIE XYZ colour space [2]. By clustering over a 1-dimensional perceptual space, rather than a linear space of 3-dimensions (or more in the case of spectral rendering) both the complexity and storage requirements of the clustering step are greatly reduced.

Figure 6.6 shows an example of the resulting clusters for a pixel neighbourhood around a shadow penumbra. It can be observed that the y-axis separation and the gradients of clusters can indicate the behaviour of the arriving illumination. The contributions of the three light sources of the scene can be clearly identified in the direct lighting layer ($v = 1$), and outliers in the first indirect lighting layer ($v = 2$) are evident.

Each individual cluster in the model stores aggregated statistics for the subset of samples in the respective domain:

```
ClusterStatistics{
  float[3]          R̂_v
  float[3]          Î_v
  float             min
  float             max
  integer           N : 31 bits
                    rejected : 1 bit
}
```

where $\hat{R}_v$ and $\hat{I}_v$ are the summations of the R_v and I_v values for each of the N contributions in the cluster.

Alongside these, the minimum and maximum extents of the cluster in the 1D luminance space are stored. The number of samples N making up the cluster is stored to allow R_v and I_v to be normalised, and to provide a means to compare pixel integrals across the image. A single bit from the sample count N is used to indicate whether this cluster has been rejected during the high intensity noise removal step outlined in the next section, 6.3.1.

[2] This is the same principle applied to Russian roulette path termination based on path throughput, reducing the visible effects of the increase in variance due to path termination while still improving efficiency

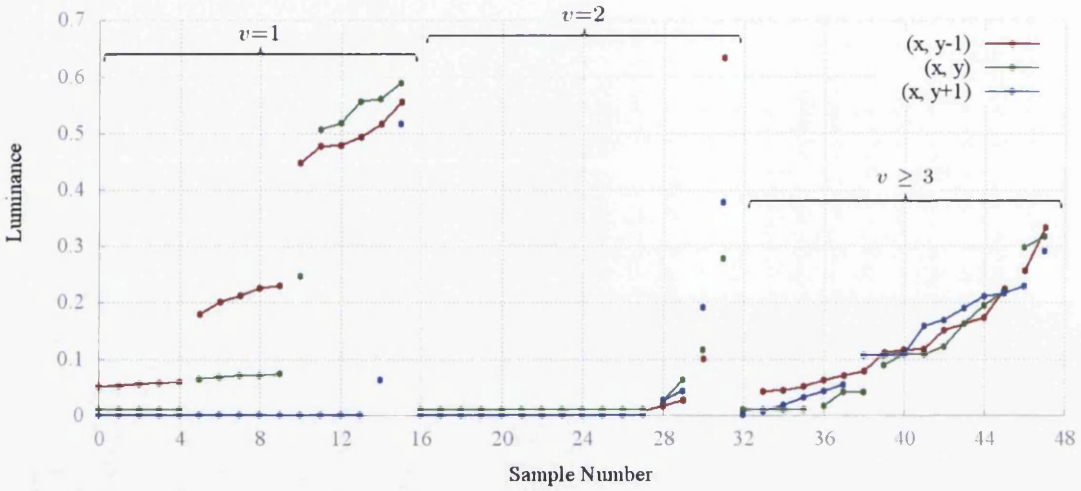


Figure 6.6: *Example plot of the layered clustering for neighbouring pixels in a shadow penumbra (see coloured points in Figure 6.5). The green and red plots are offset on the y-axis for clarity of presentation. Illumination discontinuities form separate clusters, while smooth regions are clustered together. The scene has three area light sources, whose contributions are depicted as discontinuities on the y-axis, and sample variation across the area light source by more gradual gradients.*

The aim of holding these statistics for each cluster is to enable each set of clusters to be updated dynamically, whilst respecting the quality of the integral approximation, the performance of both the clustering and filtering stages and the progressive nature of path tracing. Thus, two operations are required:

1. The addition of new contributions to a set of existing clusters,

2. The merging of existing clusters to improve performance, without sacrificing the ability to identify illumination changes.

Two heuristics are proposed that are responsible for updating and maintaining each independent set of pixel clusters for each layer based on the changes in the local integrand.

The first heuristic allows the addition of new samples to existing clusters as samples are rendered by the path tracer, creating new clusters if necessary to preserve the discontinuities in the pixel integrand. If a new vertex contribution lies within the bounds of an existing cluster or its distance from the nearest cluster is within a percentage threshold μ of the contribution, then the cluster's contribution count, N, is incremented. The R_v and I_v values for the contribution are added to the aggregated values of the cluster, $\hat{R}_v$ and $\hat{I}_v$ respectively. If the contribution lies outside the extended bounds of the existing clusters, it is assumed that the pixel integral over the luminance range of the contribution is not sufficiently represented, and a single element cluster is inserted. As a result, the parameter μ dictates how sensitive the overall framework is

to changes in luminance, and is the minimum percentage change in sample luminance that it can differentiate between.

As the clustering procedure operates alongside a progressive path tracer, it needs to respond to the changes in the integrand that come about as the estimator becomes more refined, and the distribution of contributions in the luminance domain changes. The range of luminance values covered by each set of clusters can only increase during sampling, as more of the integrand is explored, and thus discontinuities between clusters formed at the start of sampling can become less important to the overall image. To avoid a gradual increase in the number of clusters, a second heuristic is introduced that is responsible for merging existing clusters, maintaining a compact approximation of the integral.

The clustering model should be collapsed in smooth regions to reduce storage costs and computation where no discontinuities are present. This allows further refinement in regions of the integrand with rapid luminance changes, where discontinuities are more likely, and a finer representation of the sample luminance is desirable.

Based on the observations from plotting irradiance contributions (Figure 6.6), the cluster merging heuristic makes use of both the absolute luminance difference (1^{st} derivative), and the change in gradient (2^{nd} derivative) between neighbouring clusters:

$$M_{(i,i+1)} = (\Delta C_i - \Delta C_{i+1}) \cdot \phi + \frac{min_{C_{i+1}} - max_{C_i}}{max_{C_n}} \tag{6.7}$$

where ΔC is the gradient of the cluster C:

$$\Delta C = \frac{max_C - min_C}{N_C - 1}$$

To perform cluster merging $M_{(i,i+1)}$ is evaluated for each neighbouring cluster pair in the current layer, combining the statistics of clusters where M is smallest in order to maintain a minimal cluster set:

$$\begin{aligned} min_C &= Min\{\, min_{C_i},\, min_{C_j} \,\}, \\ max_C &= Max\{\, max_{C_i},\, max_{C_j} \,\}, \\ N_C &= N_{C_i} + N_{C_j}, \\ R_C &= R_{C_i} + R_{C_j}, \\ I_C &= I_{C_i} + I_{C_j}. \end{aligned}$$

Clusters for each layer are stored in ascending order of luminance to improve performance whilst processing and amending the clusters. A constant ϕ is employed to weight the influence of gradient change in M, and a value of 0.75 has been found to behave well by experimentation.

There are a number of improvements brought about by inclusion of the cluster gradient into the heuristic, that cannot be handled by relying on absolute change alone. Over time, through the progressive extension of clusters by new samples, the *min* and *max* luminance ranges of clusters in the same pixel can become similar, even if they are a result of differing illumination features. However, such continuities may instead be present as discontinuities in neighbouring pixels, if one of the features is not present. Such occurrences are useful to

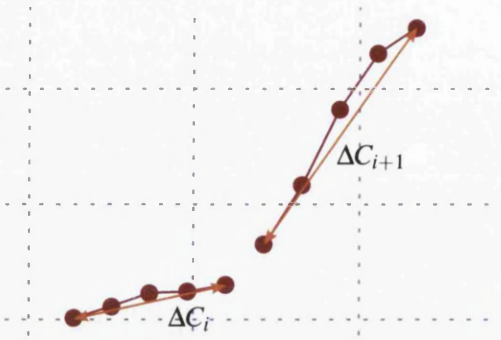


Figure 6.7: *Incorporating the 2^{nd} derivative luminance gradient into the cluster merging heuristic allows distinctive features of the illumination to be preserved, despite their similarities in absolute luminance.*

identify reduced correlation in image based filtering. Incorporating the gradient change allows patterns of similar luminance to be preserved even if they come to occupy similar space in the 1D luminance domain, based on their differing gradient characteristics (see Figure 6.7).

Favouring gradient change over absolute luminance using ϕ provides a better balance between cluster merging at extreme ends of the luminance range. For low intensity samples the gradient change is comparatively small, so increasing ϕ favours merging between lower energy clusters, where discontinuities have less visible impact. For high luminance samples the opposite is true, and high values of ϕ separate high intensity outliers and discontinuities.

This second heuristic has more impact with respect to filtering performance and quality. The framework is not overly sensitive to the initial parameter μ and values between 0.1 and 1.0 ($\pm 10\%$ to $\pm 100\%$) produce visibly similar results for the test scenes in Section 6.4. The same μ value of 0.5 is used for all images rendered using the work in this chapter.

During the early stages of the process, μ determines the number of initial clusters that are formed. A small value of μ increases the likelihood of creating new clusters from each subsequent sample, however the total number of clusters is still constrained by the less greedy cluster merging metric. Increasing μ results in fewer clusters being formed which can improve performance for slowly varying integrals, but can miss discontinuities if μ is too large. To put an upper bound on memory usage, the number of clusters stored for any one pixel layer is constrained via cluster merging. In the implementation described here a limit of eight clusters per layer is used, deemed empirically sufficient for the scenes provided here (memory usage and cluster statistics are discussed and presented in Section 6.4). Only enough clusters are needed to identify the visible changes in that particular layer of the illumination.

Paths are rendered and their contributions clustered in batches of four per pixel, reducing the overhead of cluster operations, while maintaining the short frame updates desirable in progressive rendering. In the initial pass, each contribution forms a cluster, since there is little information available about the distribution of samples in the integrand.

6.3 Noise removal and filtering

The novel clustering framework detailed thus far has a number of useful traits regarding the separation and identification of illumination features. In turn, these properties permit the development of two complimentary noise reduction techniques, geared towards preserving illumination features. First, high intensity noise removal is tailored towards temporarily removing

samples that are the cause of the white speckles typical in Monte Carlo rendered images of scenes containing specular materials. The samples responsible for this effect are generated with low probabilities, thus their presence at the start of the rendering process means that their contribution is disproportionate to the number of samples rendered so far. Until the sample distribution is normalised (according to the law of large numbers, recall Section 3.2) these samples are typically problematic due to their high energy. As a result of the clustering procedure above, these outliers can be removed efficiently and conservatively using density estimation of local clusters, discussed in Section 6.3.1. However, removal of these samples does not reduce the general noise caused by Monte Carlo estimator variance. Thus, the second filtering technique, discussed in Section 6.3.2, addresses the general noise of the estimator. Once again the clusters produced in the previous section indicate where discontinuities lie which should be preserved. The sample statistics for each cluster enable a probabilistic approach to convolution, producing a set of kernel filter weights as input for a cross-bilateral filtering scheme.

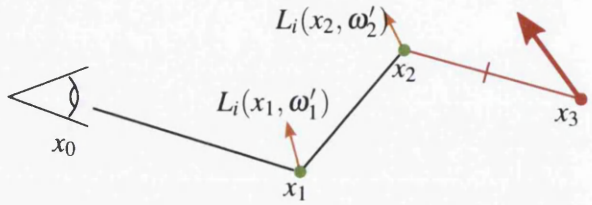
6.3.1 High intensity sample rejection

The high intensity noise removal approach proposed in this section builds upon recent density estimation methods by DeCoro et al.[DWR10] and Pajot et al.[PBP11] discussed in Section 4.4.4. The premise behind such methods is to rely on density estimation over a joint domain consisting of colour and Euclidean spaces to identify statistically outlying samples that do not conform to the current distribution for a given pixel or pixel neighbourhood. These *outliers* are then temporarily withheld from the image to avoid introducing spiked noise, and introduced individually as they become statistically similar to the ever changing pixel distribution of the progressive renderer.

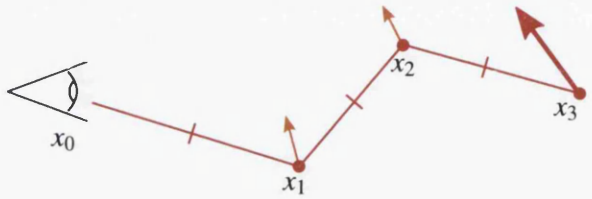
The approach taken here aims to extend these principles, improving on efficiency and temporarily incurred bias. In addition there is a focus on the suitability of such techniques as a pre-process to image-based filtering. Building upon the newly introduced clustering framework, affords a number of benefits to the approach presented here, and a number of key differences from previous works:

- Per vertex contributions are considered individually, as opposed to composited into the final radiance of a complete path,
- Intermediate incident radiance values (the I_v values from the clustering framework) form the basis of density estimation and contribution rejection, instead of the final radiance measurements,
- More lenient but effective noise removal is applied, suitable at low sample counts and for difficult to sample illumination.
- Outlier rejection can be adjusted post-render, utilising the clustered data to fine tune the noise removal to achieve the desired results.

Layer based sample rejection Relying on composite path radiance means that previous approaches cannot isolate the cause of high variance and noise within a path sample, and hence



(a) Layer based outlier rejection



(a) Composite, path-based outlier rejection

Figure 6.8: *Layer-based noise removal (a) can isolate high energy contributions (x_3) and retain vertex contributions that conform to their respective path space distributions (x_1 and x_2), which are otherwise rejected by previous composite approaches (b).*

must remove the contributions from all vertices of the path (Figure 6.8b). Thus, despite only a single vertex being responsible for the increased pixel variance, contributions that can potentially improve the estimator for the remaining path space are also removed. This results in reduced efficiency and introduces avoidable bias to the estimator. The number of paths temporarily removed from the estimator is increased, subsequently removing energy from the image introducing noticeable darkening of the image.

Conversely, the technique presented here utilises per-vertex contributions as described in the clustering process presented above, allowing isolation of the illumination effects and hence also the noise contributed at each layer by the respective samples. As mentioned in the motivation for this work at the beginning of the chapter, the variance of the estimator increases on average in higher dimensions due to the variation in sample density at each vertex. High energy noise at a single vertex can therefore be removed whilst the remaining vertices still contribute to the estimators for their respective path lengths, illustrated in Figure 6.8a).

Texture independent density estimates The second key difference is that outliers are processed with respect to their incident radiance, as opposed to the final pixel radiance they contribute to the image. Working in the space of incident radiance has be done before [KRK04], but not applied to outlier rejection techniques. In previous approaches, relying on final radiance values means that the ability to reduce noise, caused by the high variance illumination, is dependent on the texture changes across the 2D image neighbourhood. This presents two problems:

1. Complex texture detail increases the colour space distance between samples, reducing the local density.

2. High frequency edges can increase the local density, preserving unwanted noise.

These are somewhat inverse outcomes of the same fundamental problem, where the colour space distance of contributions is artificially increased and decreased respectively by texture

detail. This can be seen as a form of boundary bias (recall Section 4.2 regarding bias in photon mapping) but as opposed to just geometric edges, changes in texture detail also form discontinuities across the filter kernel. Over a uniform region of a textured surface, the density of contributions for a given region of the image is reliant solely on the underlying incident radiance at the surface, producing differences in colour space. Thus, outlying noisy samples will be removed effectively. Inverting this assumption, such that the incident radiance is constant across image space and the texture detail varies, means that intuitively no samples should be removed since the radiance is uniform. Variation in the pixel radiance is now only a result of the texture detail, which is of low variance providing good sampling techniques and adequate texture filtering has been applied. However, changes in texture detail are now solely responsible for the local density change in colour space and subsequently, contributions incident on regions of fine or high contrast texture detail may be removed. By eliminating the effect of texture maps on density estimates and relying on incident radiance the technique described here provides a more intuitive approach to removal of high intensity contributions; Only dimensions of the integrand where the majority of the variance is introduced into the estimator are included in the density estimation.

Effects of the filtering parameters As well as the luminance similarity measure μ introduced previously which represents the colour space distance between contributions, the high intensity noise removal relies on two additional parameters:

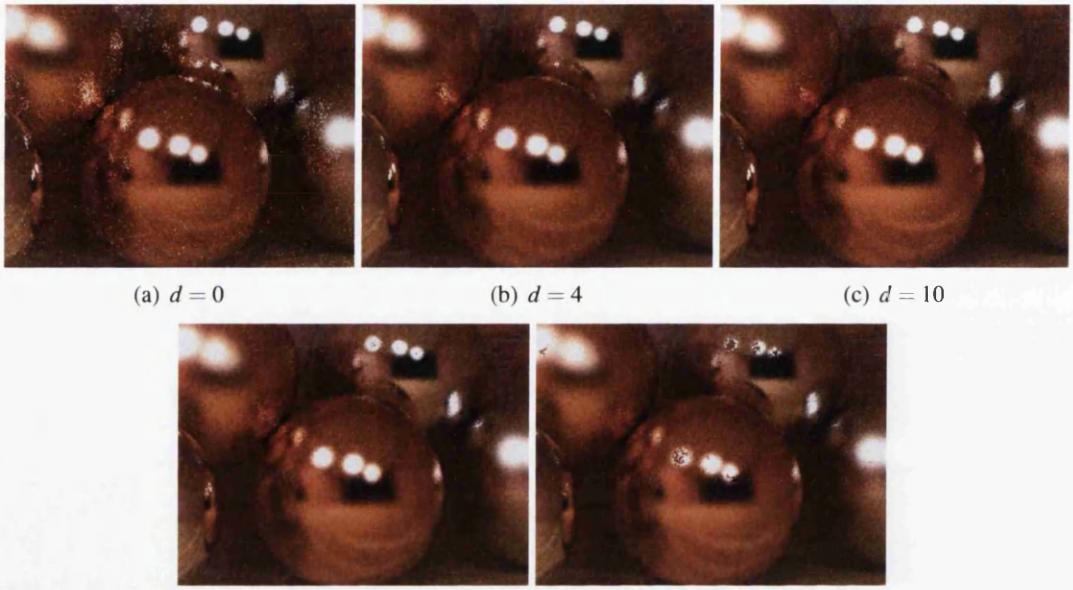
- σ, specifies the filter kernel bandwidth defining a neighbourhood in 2D image space,
- d, specifies the global density threshold,

The global density threshold d dictates the minimum number of neighbours n within the neighbourhood defined by μ and σ, below which a contribution is rejected:

$$d = \frac{n}{\sigma^2 \mu}$$

In the approach presented here, the 2D spatial and 1D luminance domains (σ and μ) are kept constant across the image, leaving the behaviour of the high intensity noise removal to d. Since σ and μ are fixed, the density d can be normalised and thus n and d are proportional and used synonymously.

Figure 6.9 shows the effect of adjusting the global density threshold d for a set of contributions in a scene with glossy and specular metallic BRDFs. As the density threshold increases for a fixed set of contributions, more are rejected from the final image. If $d(n)$ is too high, the rejection strategy becomes overly harsh and effects become noticeable, especially at high contrast edges and gradients. Even low values of n provide a noticeable improvement in the objectionable noise, but using this technique alone does not remove noise in its entirety. Not only is the aim to remove persistent high intensity noise, but to reduce all high frequency noise visible in the image.



(a) $d = 0$ (b) $d = 4$ (c) $d = 10$

(d) $d = 20$ (e) $d = 50$

Figure 6.9: *Effects of changing the density threshold parameter, d, of the high intensity noise removal with* $\sigma = 11$ *and* $\mu = 0.5$ *as used for the Glossy spheres scene. Timings can be found in Figure 6.10.*

Energy preservation at low densities Convolution related filtering techniques reliant on weighting a neighbourhood of samples are highly influenced by high intensity noise, but excel when faced with more subtle noise where the energy of individual input values is more uniform. Thus they are a logical choice in partnership with density based noise removal techniques. In addition to the remaining noise left behind, a drawback of the techniques proposed by DeCoro et al.[DWR10] and Pajot et al.[PBP11] is that their rejection principles are not immediately suitable for use alongside such image-based convolution filters.

Neither method operates well under low sample densities, resulting in a large proportion of contributions being removed across a region, providing few input samples for a subsequent filter. In the work proposed here, the view is taken that it is better to reject contributions leniently with a risk of artefacts, such as can be seen in Figure 6.11, than to apply an overly aggressive strategy and remove significant radiance entirely from a region. Doing so would provide no indication of the underlying materials or light transport for those pixels, and furthermore remove the ability of any successive filter to improve the perceptual image quality.

6.3.1.1 Noise removal procedure

Algorithm 4 provides an outline of the noise removal strategy. Computing density estimates for all clusters over a fixed kernel radius means the computation of the actual density is skipped since it is implicit by fixing σ and μ. Thus as soon as the required minimum density has been

Algorithm 4 High Intensity Noise Removal

```
1:  L_max ← 0                                              ▷ Highest accepted luminance
2:  for depth ← 0 → 3 do
3:      density ← 0                                        ▷ Num. similar samples
4:      for c ← <num. clusters in layer> → 0 do            ▷ for each cluster
5:          C ← Clusters[c]
6:          density ← density + C_num
7:          bool t ← density ≥ n                           ▷ where n is the local density threshold
8:          if C_min ≤ L_max then t ← true
9:          if !t then
10:             for all pixels P in σ² kernel do           ▷ Check pixel neighbourhood
11:                 for all clusters C_p in P do
12:                     if Overlap(C_p, C) > 0 then
13:                         density ← density + C_pnum     ▷ include similar samples from P
14:                         t ← t ⊕ (density ≥ n)          ▷ where ⊕ is boolean OR
15:                         if t then
16:                             L_max ←Max(L_max, C_max)
17:                             Clusters[c] ← set as enabled
18:         c ← c − 1
```

reached (ie: $density \geq n$, Algorithm 4, lines 7 and 15) the cluster in question can be accepted. In the work by DeCoro et al.it is n that is fixed, and a nearest neighbour search is required to obtain a value for the minimum radius enclosing the n nearest samples. By fixing the kernel radius as opposed to n, much faster searching is achievable since the spatial domain is fixed and an iterative expansion or contraction of search space is not needed (can be required in k-NN searches), in addition to the performance improvements brought about by searching in reduced dimensions. Furthermore, 2D image space is already partitioned uniformly into pixels and further partitioning is afforded by the clustering over the luminance domain, providing an opportunity for fast searches and the exclusion of distant contributions in both spatial and luminance domains. As a result of these optimisations, the kernel width σ can be extended with sub-linear performance impact (Figure 6.10). This can improve filtering for larger regions and soft gradients, whilst still removing patches of high intensity noise in regions corrupted with sparse general noise.

Using a clustering framework to represent a pixels sample distribution, as opposed to a sparse set of individual samples as used by previous approaches, requires a means of estimating the number of similar samples in the clusters of neighbouring pixels. The existing parameter μ used during clustering is employed again to determine cluster similarity in luminance space for establishing local densities. Since the clusters of each pixel are constructed independently, clusters cannot be compared directly.

Treating clusters as sub-integrals of their respective path space, the distance in luminance space between any two clusters can be computed as a proportion of their overlap in this space. In other words, the intersection of the areas A of the definite integral for the current cluster C

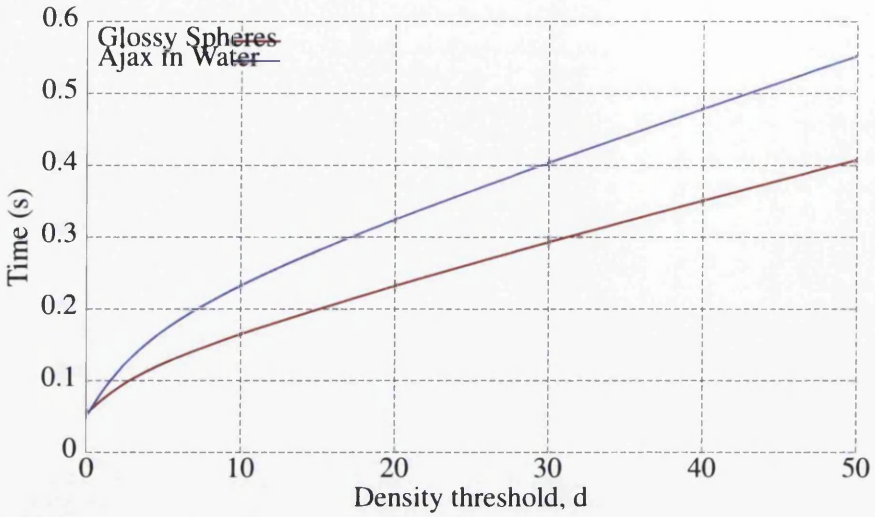


Figure 6.10: *Increasing the minimum density threshold for each contributing sample results in a sub-linear increase in computation, which is negligible when compared with the costs of rendering (see Section 6.4 for the respective images and data for these scenes).*

and a neighbouring cluster C_p are computed, as a ratio of C_p:

$$\text{Overlap}(C_p, C) = \frac{A(C') \cap A(C_p)}{A(C_p)} \tag{6.8}$$

where C' is the current pixel cluster extended according to μ. This accounts for the luminance range covered by the *min* and *max* extents of the cluster, and to clusters that are similar within the luminance domain according to μ:

$$min_{C'} = min_C \cdot (1 - \mu)$$
$$max_{C'} = max_C \cdot (1 + \mu)$$

Storing clusters in ascending order of luminance in the framework improves the performance of this stage of the algorithm. Ordering the clusters allows high luminance clusters to be processed first, and if accepted, the remaining non-zero clusters in the pixel can be enabled (Algorithm 4 line 1). This is possible since samples of lower energies can be effectively handled during the irradiance-aware filtering stage (introduced in the following section), so only contributions supplying high intensity noise are of interest for removal here. This allows the remaining clusters to be enabled easily with high efficiency, and also following the mantra of preserving energy through minimal rejection rates.

Note that L_{max} stores the maximum accepted luminance of the cluster in the current pixel, not per individual layer. The accepted cluster with luminance L_{max} will dominate the intensity of the pixel, so rejecting clusters of similar luminance will have little effect on the expected value of the pixel. Consequently, it also speeds up the density estimation as the actual local density of this contribution need not be calculated and contributions from neighbouring pixels need not be queried.

After the noise removal procedure has been run once for a given pixel, only incremental updates are required when new clusters are created. The rejection status of each cluster is persistent throughout rendering, stored in alongside the cluster statistics. When new contributions are added to an existing cluster, they are automatically accepted or rejected by association, based on that clusters current rejection status. For a new cluster, its *max* luminance is tested against the existing L_{max} value of the highest accepted cluster in the respective pixel. This can be found trivially since the clusters are ordered, and if $max < L_{max}$ then the cluster is flagged as accepted, otherwise it is rejected to avoid the potential introduction of noise (since the full neighbourhood density is not being computed). The noise removal algorithm is run again after a set number of frames or when parameters change, evaluating the full neighbourhood density where necessary on newly created clusters.

In contrast to previous approaches, the parameters d, μ and σ can be modified while rendering, or after rendering is complete. This is due to the fact that a complete, compact representation of all contributions is maintained throughout using the per-pixel clusters, thus allowing individual clusters to be added and removed from the image as desired. This permits the user to fine tune the noise removal and to clean up the image to the desired levels, while maintaining minimal bias through reduced rejection rates.

6.3.2 Illumination preserving filter

The advantage of the rejection based method above is that typical filtering artefacts in the image, generally caused by blurring or smearing of features, are not present. Bias is instead introduced by removal of existing samples, but such an approach is inadequate for reducing the residual noise. After applying the high intensity noise removal algorithm proposed in this section, the visual quality of the illumination features is improved, by introducing a novel technique built upon the principles of the cross-bilateral filter, discussed in Section 4.4.5. The general premise is that more reliable and low variance data is utilised to identify high frequency changes that need to be preserved when de-noising the target image. Depth, normal and texture buffers are typically used to identify changes in geometry (Figure 6.12) that produce low noise and are reliable with minimal sampling, as well as being readily available as a by-product of the ray tracing process.

Revisiting the Cross-Bilateral Filter Using these range buffers in conjunction with Gaussian convolution provides an effective scheme for low pass smoothing, but can struggle to reduce noise in regions where the range buffers are highly restrictive. Recall the procedure behind the CBF discussed in Section 4.4.5. To calculate the filter kernel for a given pixel p using a cross bilateral filter (CBF), the distance between pixels $q \in S$ (the 2D pixel neighbourhood) and p need to be computed for each range buffer, to provide the range terms for the Gaussian convolution. Revisiting Equation 4.12, the output value for a pixel p using a series of range buffers $RB^0, ..., RB^n$ can be computed:

$$CBF[I, RB^0, ..., RB^n]_p = I_q \frac{1}{Z_p} \sum_{q \in S} G_\sigma(\|p-q\|) G_{\sigma_r^0}(|RB_p - RB_q|) \quad ... \\ G_{\sigma_r^{n-1}}(|RB_q^{n-1} - RB_q^{n-1}|) G_{\sigma_r^n}(|RB_p^n - RB_q^n|) \tag{6.9}$$

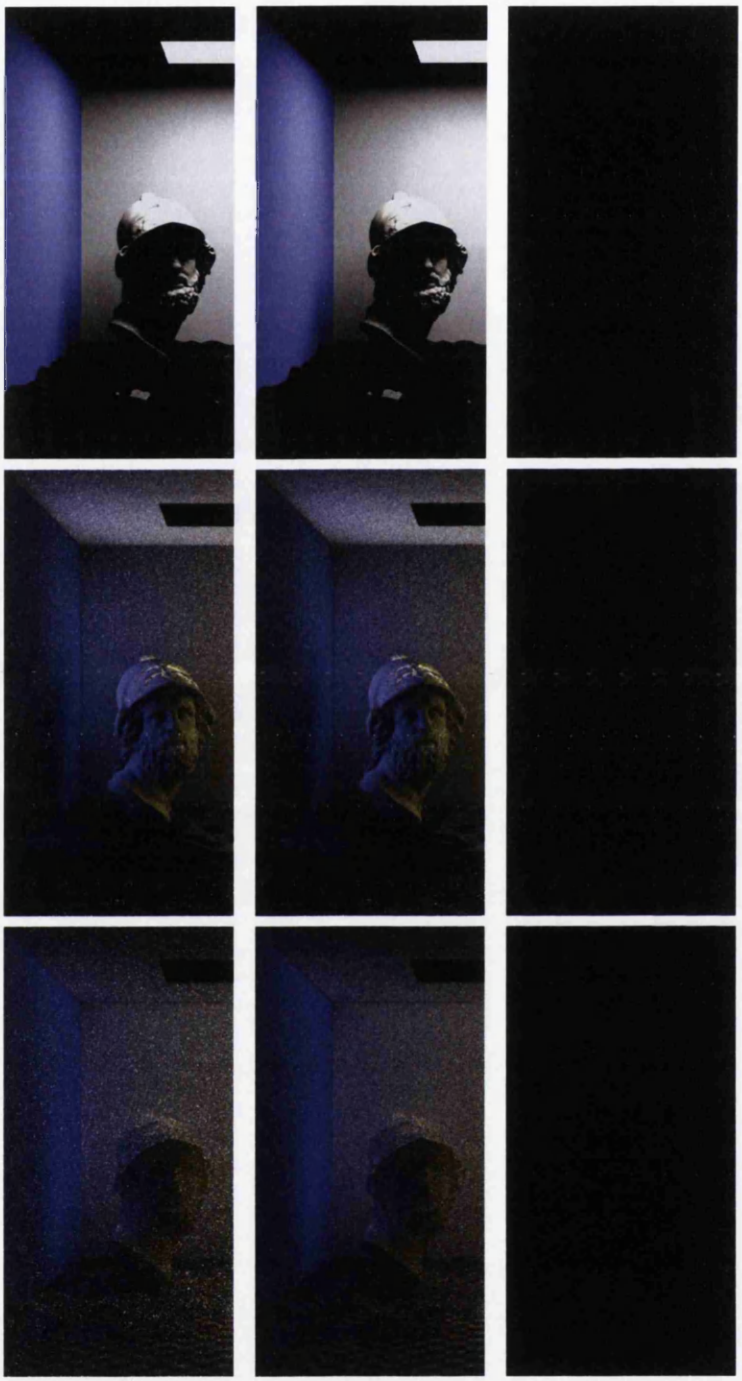
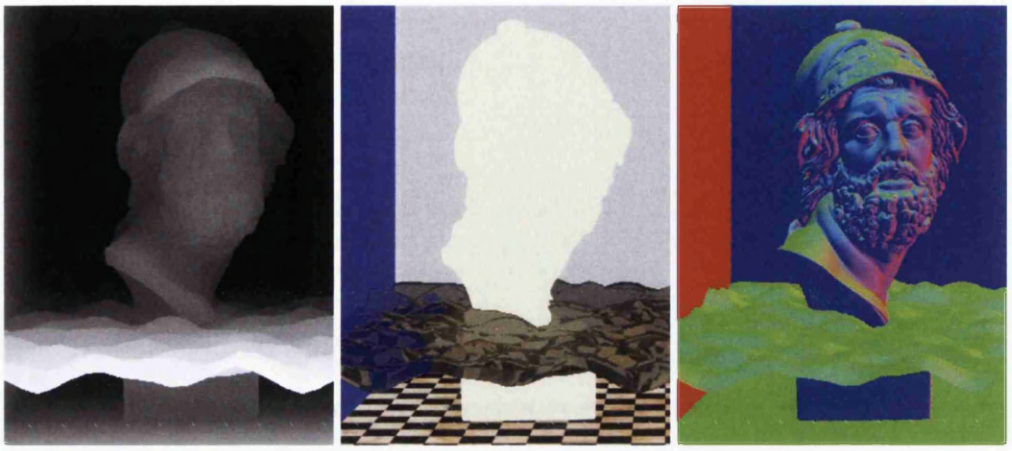


Figure 6.11: *Example of the high intensity noise reduction applied to each layer of vertex contributions from the Ajax bust scene, using the parameters described in Section 6.4. From top to bottom the layers are* $v = 1$, $v = 2$, $v \geq 3$.



(a) Depth buffer (b) Texture detail (c) Surface normals

Figure 6.12: *Low variance buffers are relied upon to identify edges in the geometry that should be preserved during filtering. Previous approaches also rely on the texture buffer to preserve edges in the image, but this can limit the effectiveness of the filter in small regions of fine detail.*

thus the final weight of a pixel q can be drastically influenced by only one of the set of n range buffers. Thus as n increases the number of influential pixels in complex environments will diminish, increasing the output qualities reliance on a small number of samples from the noisy input image. Alternatively the overlapping features identified by each filter may lose the edge preserving qualities of the convolution as a whole. In other words, the net effects of combining the n range buffers can in the worst cases produce a convolution kernel with equal weights, or have no effect on the input at all.

In addition, small features in the range buffers, such as from highly detailed textures, can significantly reduce the number of useful pixels in the kernel that are given significant weights. Intuitively, as the number of pixels in the kernel decreases, the signal to noise ratio is affected due to the relative increase in variance of the stochastic estimators. Since the weights are normalised in order to preserve energy, the influence of these few pixels on the filtered output of p increases. As a result, noise in small regions becomes amplified rather than reduced.

To combat this, and to maximise the number of influential samples in the filter kernel, the technique proposed here relies only on the depth and normal buffers to isolate geometric edges. Thus the effectiveness of the new technique is independent of texture features, resulting in more usable and influential pixel data given the same kernel size. Accomplishing this is down to the clustering framework that operates over incident radiance at each vertex, eliminating the effects of texture detail on the pixel intensities allowing those alone to be filtered and the texture detail preserved.

Preserving illumination features Relying exclusively on low variance data means that features not present in the range buffers cannot be preserved in the filtered image. The source of the noise being targeted by filtering is in the high variance illumination, thus any discontinu-

ities present in the radiance field are not taken into account when applying the low pass filter. This can present significant artefacts around high frequency edges as a result of blurring hard shadows, and not maintaining specular and glossy reflections, as will be seen when discussing the results of the approach proposed here.

Unlike the geometry, the incident radiance is made up of layers each of which have their own features and phenomena that should be preserved in the final image. Following the approach taken with clustering and the high intensity noise removal algorithm, the novel low pass filter also operates over each layer of the incident illumination individually, with attention to only the discontinuities present in the current layer. By doing so, the corruption due to noise from features in the remaining layers is eliminated improving the signal to noise ratio with respect to the discontinuities making them easier to identify.

Again, as with the removal of the texture based range buffer, this strives to minimise the complexity of the range buffer terms, maximising the number of pixels that influence the resulting pixel value whilst adhering to only discontinuities relevant to those illumination features. This allows indirect lighting to be filtered across boundaries in the direct illumination, boundaries in other layers of the incident illumination and vice versa. This provides benefits that are not achievable using a single layered approach.

To handle these discontinuities and boundaries in the illumination, commonly caused by light source occlusion and high frequency BSDFs which bring both strong edges and high variance to the image, the discrete Poisson probability distribution is employed. This provides additional filter weights that detect visible changes in the high variance signal, and not present in the geometry based range buffers.

6.3.2.1 Poisson-based filter weights

Due to the central limit theorem [3], the sample mean for a Monte Carlo estimator will have an approximately normal distribution, for N independently drawn samples. This property was used by Kontkanen et al.[KRK04] under the assumptions of slowly varying, low frequency illumination to provide a model for statistically estimating irradiance for a pixel. However, it is unknown when (ie: for what N) the distribution of such pixels will approach a normal distribution, and such an approach is not feasible when assumptions on the frequency of the illumination are lifted. Instead, the approach taken here is to analyse the occurrence of samples in a specific luminance band across a pixel neighbourhood; thus providing a statistical model of a distribution across pixels, as opposed to across the luminance range of a single pixel.

During rendering, given two independent distributions of N samples from arbitrary integrands (ie: a set of samples from two pixels), it would be desirable to establish whether or not they follow the same distribution, and hence determine their similarity with respect to the current features of interest in the integrand.

However, for low N, the central limit theorem and hence the assumption of a normal distribution does not apply. As a result, the Poisson distribution is turned to, as it can be applied to rare events and provides more reliable estimates at low sample counts, allowing an estimate of

[3]The central limit theorum states that for a given distribution, the mean of a large number of independent random variables will be close to normally distributed

integrand significance to still be attained. Recall from Equation 6.3 that for a Poisson random variable X the probability that it equates to k is given by:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

In other words, what is the likelihood that given an average occurrence of λ events in a given interval, k events will be observed in any other interval under the same process(es)? It is obvious that given two pixels p and q the total number of contributing (ie: non-zero) samples alone cannot be relied upon to provide a similarity measure, since the variation in pixel intensities is dominated by the values taken on by each of the random samples, not the occurrence or absence of the samples themselves. Therefore, in order to obtain a statistical model that is meaningful with respect to the final pixel radiance, the luminance or intensity of the samples must be taken into account, in addition to their occurrences.

Assume first, that there exist two pixels with identical integrals, such that the expected values of their estimators are equal in the limit, ie: $\langle I_p \rangle_N = \langle I_q \rangle_N$ for large N. After N samples, the distributions of their respective samples over the integrand should be similar to one another, with respect to the normal distribution. Similarly, taking any respective part of the two integrals, such as samples within a specific luminance range, will yield similar properties with respect to the sample distribution.

Taking these two pixels, assume that for a given luminance range a number of samples N_p and N_q have been observed from pixels p and q respectively. As a result the expectation would be that the probability of observing N_p events in the estimator of q would be such that:

$$P(N_q = N_p) = \frac{N_q^{N_p} e^{-N_q}}{N_p!} \tag{6.10}$$

is maximised, having substituted $\lambda = N_q$ and $k = N_p$ in the Poisson approximation. This is based on the fact that both N_p and N_q have been observed over the same interval, resulting from the integrals of p and q being sampled equally via the uniform distribution of samples across pixels in the image. In the case of MC path tracing this is true, and an extension to adaptive rendering for non-uniform distributions over the image is discussed later in the context of future work.

If the assumption about the similarities of the integrals of p and q is removed, then the probability $P(N_q = N_p)$ can be expected to reduce in relation to the similarity of the pixel integrals according to the Poisson distribution. In the context of filtering, this statistical similarity can be used to compute filter kernel weights for each pixel q in the kernel, centred around the target pixel p that is being de-noised. For a pair of pixels where $P(N_q = N_p)$ is large the pixels are correlated and thus q should have influence on the filtered value of p. Small or zero values for $P(N_q = N_p)$ indicate highly varied distributions, and a discontinuity in the luminance integral and so the weighting of q on p should be minimal.

Formation of kernel groups Having built a sample clustering framework around luminance bands means that for a given pixel, the distribution of samples with respect to luminance is

Algorithm 5 Formation of kernel groups

```
1:  KGroups ← initialise kernel groups from clusters in p
2:  for all pixels q ∈ σ² do                          ▷ Each pixel in the filter kernel
3:      for all clusters C_0 to C_n in q do
4:          for all groups K in KGroups do
5:              b ← Overlap(C_i, K)                    ▷ Equation 6.8
6:              if b > 0 then
7:                  add statistics of C_i to K         ▷ Equation 6.11
8:                  if min_K or max_K have changed then
9:                      Check and merge existing KGroups
10:                 break
11:         if b = 0 then
12:             insert C_i into KGroups                ▷ Create a new kernel group
```

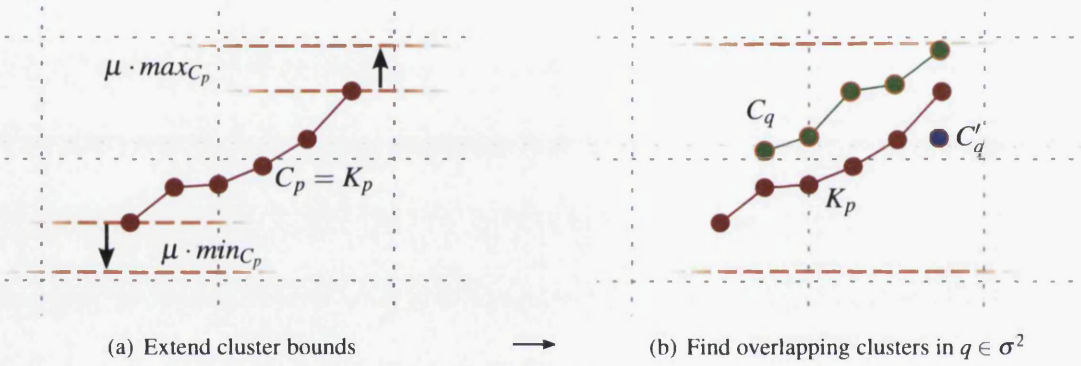


Figure 6.13: *Creation of a kernel group involves extending the bounds of existing clusters in the target pixel (according to Equation 6.8)(a), and computing the overlap with neighbouring pixel clusters for $q \in \sigma^2$ (b)*

preserved and as a result the ideas above can be implemented in the clustering framework describe thus far. An overview of this process is shown in Algorithm 5.

As was the case for the high intensity noise removal, dealing with clusters requires a mechanism for identifying similar clusters across a set of pixels. Using the same parameter μ as before, the overlap between the clusters can be computed following Equation 6.8 to compare any two clusters. However, for convolution based filtering the relative similarities between all pixels in the kernel are required, in order to construct weights for the filter kernel. Therefore an extension to this mechanism is required, identifying similarity between clusters from multiple pixels, so that their influence on the de-noising process can be adjusted accordingly, via normalisation over the kernel. To this end, a top level clustering scheme is introduced that temporarily groups existing clusters from different pixels q in the kernel S (determined by the parameter σ from before) based on their luminance range (as described by Algorithm 5).

Despite being $O\left(|C| \cdot |K| \cdot \sigma^2\right)$, where $|C|$ is the size of the set of clusters C the number of clusters per pixel is constrained by the framework such that $|C| \leq 8$ and for most cases $|K| \lesssim |C|$

due to the correlation between pixel integrals (note that if such correlation was not present then low pass convolution filters would not be of use in the first place). For slowly varying or low frequency illumination, which is prevalent in the real world and hence also in photo-realistic imagery, a single kernel group is produced representing all samples across a neighbourhood. In more complex regions that includes illumination edges, multiple kernel groups are produced each of which represents a distinct band of luminance. Together these groups aim to represent the visible discontinuities of the incident radiance within the filter kernel, much like the role the individual clusters hold for each pixel.

Each kernel group K is responsible for keeping track of the contributions from clusters of each pixel for the luminance band covered by K:

```
KernelGroup{
        float           min     Min luminance
        float           max     Max luminance
        integer[σ²]     N̂       Sample counts
        colour[σ²]      R̂       Radiance contributions
        colour[σ²]      Î       Irradiance contributions
}
```

Initially, this set of kernel groups is populated using the clusters in the centre pixel p that is being de-noised, by simply creating a new kernel group from each pixel cluster. For subsequent pixels, a cluster C from q will commonly overlap with an existing kernel group K, due to the presence of pixel correlation over the filter kernel, and the data in K for q must be updated from C (algo. 5 line 7):

$$\begin{aligned} min &= Min\{\, min,\ min_C \,\} \\ max &= Max\{\, max,\ max_C \,\} \\ \hat{N}_q &= \hat{N}_q + N_C \\ \hat{R}_q &= \hat{R}_q + R_C \\ \hat{I}_q &= \hat{I}_q + I_C \end{aligned} \tag{6.11}$$

hence for a given set of clusters $C_0, \ldots, C_n \in q$, the resulting kernel group statistics are computed as:

$$\hat{N}_q = \sum_{i=0}^{i=n} N_{C_i} \quad [C_i \cap K \neq 0]$$

$$\hat{R}_q = \sum_{i=0}^{i=n} R_{C_i} \quad [C_i \cap K \neq 0]$$

$$\hat{I}_q = \sum_{i=0}^{i=n} I_{C_i} \quad [C_i \cap K \neq 0]$$

where:

$$C \cap K = \begin{cases} 0 & \text{if } max_C \leq min_K \text{ or } min_C \geq max_K, \\ 1 & \text{otherwise} \end{cases}$$

in other words, for all clusters C that overlap K in the luminance domain. As a result, if the bounds of K are extended, it may be possible to merge existing kernel groups (Algorithm 5 line 9). This merging process maintains a minimum set of kernel groups, reducing work required to compute filtering weights using the Poisson approximation.

Filter weights and final radiance calculations are performed for each kernel group independently, allowing multiple overlapping illumination sources to be de-noised. The filter weights for each kernel group are calculated from the statistics of each contributing pixel q (where $C \cap K \neq 0$), using Equation 6.10. These Poisson weights, denoted W_λ are combined with a subset of the more conventional weights used in cross-bilateral filters, namely:

1. $W_{\sigma_g^2}$, from the Gaussian distribution over σ^2,
2. W_{σ_d}, based on the depth values for each pixel,
3. W_{σ_n}, from the surface normal orientations.

The resulting weights are a product of these sets of filter weights for the kernel cluster:

$$W_p^K = W_{\sigma^2{}_g}(p) \cdot W_{\sigma_n}(p) \cdot W_{\sigma_d}(p) \cdot W_\lambda(p) \tag{6.12}$$

All weights excluding W_λ can be computed once per pixel since the underlying geometry is constant across all samples. The resulting convolution kernel, weights the influence of each contributing pixel while adhering to both geometric changes and features in the luminance of the incident illumination field across the kernel. Using this filter kernel, the de-noised final radiance value for a pixel can be calculated from the data in the kernel groups.

6.3.3 Pixel radiance computation

Each kernel group K stores $\hat{N}$, $\hat{R}$ and $\hat{I}$ as a sum of corresponding values in the per-pixel clusters that overlap K. To remove noise from the centre pixel p it is desirable to combine the radiance $\hat{R}$ in the neighbourhood to remove noise but in doing so, minimise the bias introduced to p. This achieved by isolating components of $\hat{I}$ and $\hat{R}$ from p to preserve high frequency changes like texture detail, whilst convoluting the remaining components to reduce noise in the incident radiance. First, recall the calculations for the radiance and incident radiance for each path vertex contribution v, from Equations 6.4 and 6.5:

$$R_v = L_i(x_v, \omega_v') \prod_{i=1}^{v} f_r(x_i, \omega_i, \omega_i')(\mathbf{n}_i \cdot \omega_i') \tag{6.4 revisited}$$

$$I_v = (\mathbf{n}_d \cdot \omega_d') L_i(x_v, \omega_v') \prod_{i=d+1}^{v} f_r(x_i, \omega_i, \omega_i')(\mathbf{n}_i \cdot \omega_i') \tag{6.5 revisited}$$

and since these are products of many terms, the omitted terms from I_v can be found using Equation 6.6:

$$\frac{R_v}{I_v} = f_r(x_d, \omega_d, \omega_d') \prod_{i=1}^{d-1} f_r(x_i, \omega_i, \omega_i')(\mathbf{n}_i \cdot \omega_i') \tag{6.6 revisited}$$

From a practical perspective, a BRDF $f_r(x_i, \omega_i, \omega_i')$ is composed of two distinct elements, as discussed in Chapter 2. The scalar scattering coefficient, $f^s(x_i)\omega_i\omega_i' \in (0,1]$, dictates the proportion of light from ω_i scattered towards ω_i'. The spectral coefficients of each colour channel for the BRDF are accounted for by the texture detail coefficient $f^t(x_i)$. This describes the relative change between colour channels, and is independent of ω_i and ω_i', changing only with the position of x_i on the surface.

$$f_r(x_i, \omega_i, \omega_i') = f^s(x_i)\omega_i\omega_i' f^t(x_i)$$

Separating magnitude change ($f^s()$) and component-wise change ($f^t()$) in this way is useful for preserving local and rapidly changing per-pixel detail, whilst sharing the global incident illumination information across neighbouring pixels.

For each pixel, the $f^t(x_i)$ term for $1 \leq i \leq d$ of each Monte Carlo path is scaled by the path throughput from x_0 to x_d added to the layered detail buffer $\hat{D}$. This becomes refined during each rendering iteration such that $\hat{D}$ is the weighted sum of all $f^t(x_d)$:

$$\hat{D} = f^t(x_d)\prod_{i=1}^{d-1} f^s(x_i)\omega_i\omega_i'(n_i \cdot \omega_i')$$

One such detail buffer is stored per layer (Figure 6.14) in order to account for detail maps on glossy and specular surfaces, and to ensure the textures visible from glossy surfaces do not affect direct lighting estimates. The only remaining unknown value not present in the clustering framework is $f_s(x_i, \omega_i, \omega_i')$ which can easily be obtained by rearranging Equation 6.6 and substituting $\hat{D}$:

$$f^s(x_d) = \frac{R_v}{I_v\hat{D}_v} \tag{6.13}$$

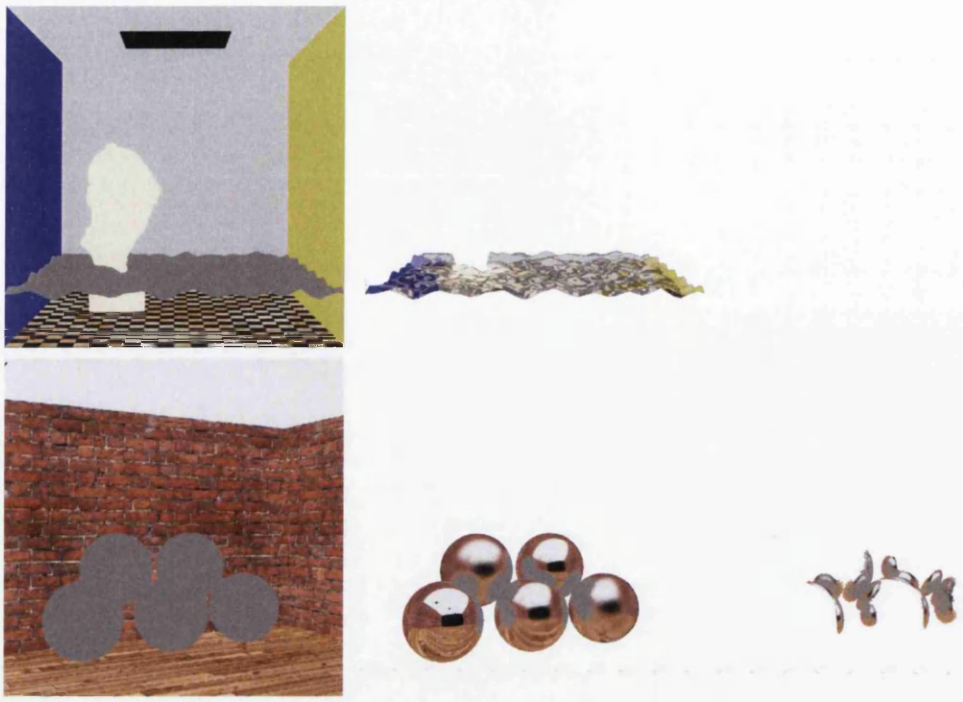
From these Equations it can be seen that the high variance radiance R_v and incident radiance I_v can be isolated from the lower frequency terms of the specular path throughput and texture detail $\hat{D}$. To leverage the power of image based convolution filters, contributions from the corresponding layer v, in neighbouring pixels are utilised to improve the signal to noise ratio of I_v, while preserving the $\hat{D}$ term for the target pixel.

Kernel group radiance To be of use in the clustering framework, this isolation of terms must be applied to the contributions represented by a cluster, and hence within a kernel group spanning the pixel neighbourhood.

From Equation 6.13, a BRDF scalar coefficient can be calculated for all contributions in a cluster:

$$\hat{f}^s(d) = \frac{\hat{R}}{\hat{I}\cdot\hat{D}}$$

representing the overall magnitude change between $\hat{R}$ and $\hat{I}$ of a cluster. To remove noise caused by insufficient sampling of the illumination, the incident radiance data $\hat{I}_q$ and $\hat{f}^s(q)$ from each contribition pixel in the filter kernel (ie: those clusters $C \in q$ where $C \cup K \neq 0$) needs to be weighted according to $W_p^K(q)$, while preserving the texture detail $\hat{D}^p$ for the pixel



$v = 1$ $v = 2$ $v \geq 3$

Figure 6.14: *The texture detail buffers $\hat{D}_v$ for each layer of the framework, representing the path throughput and texture detail buffer for each layer in the Ajax bust and Glossy spheres test scenes. Note that the Ajax scene has only two layers since specular inter-reflection does not occur.*

being de-noised. The final radiance contributed to p by clusters in pixel q for kernel group K is computed using:

$$K[\hat{R}_{p \leftarrow q}] = K[\hat{I}_q] \cdot K[\hat{f}^s(q)] \cdot \hat{D}_p \tag{6.14}$$

$\hat{R}_{p \leftarrow q}$ is then combined from each pixel using the respective filter weights from W_p^K to influence the contribution of q to the filtered radiance value of the kernel group:

$$\hat{R}_p = \sum_{q=0}^{\sigma^2} K[\hat{R}_{p \leftarrow q}] \cdot W_p^K(q) \tag{6.15}$$

providing a filtered contribution from each kernel group n, representing the filtered contributions of samples in the pixel integral at that luminance range covered by K. The total measured pixel radiance is the sum of $\hat{R}_p$ for each kernel group, normalised using the sum of the sample

Algorithm 6 Incident radiance filtering

1: $M_p \leftarrow 0$ ▷ Final radiance measurement for pixel p
2: **for all** clusters K in kernel groups **do**
3: **for all** pixels $q \in \sigma^2$ **do** ▷ Each pixel q in the filter kernel
4: $W_\lambda(q) \leftarrow \text{Poisson}(K[N_q], K[N_p])$
5: $W_p^K(q) \leftarrow W_\lambda(q) \cdot W_{\sigma_g}(q) \cdot W_{\sigma_n}(q)$ ▷ Equation 6.12
6: $\hat{R}_{p \leftarrow q} \leftarrow$ radiance from K ▷ Equation 6.14
7: $\hat{R}_p \leftarrow \hat{R}_p + \hat{R}_{p \leftarrow q} / K[N_q] \cdot W_p^K(q)$
8: $M_p \leftarrow M_p + \hat{R}_p$ ▷ Add normalised radiance

counts in those clusters:

$$M_p = \frac{\sum_{n=0}^{|K|} K_n[\hat{R}_{p \leftarrow q}]}{\sum_{n=0}^{|K|} K_n[\hat{N}]}$$

producing the final filtered radiance value for the pixel.

To improve performance, kernel groups with uniform Poisson weights for each pixel in σ^2 can be merged, reducing the number of iterations over the kernel groups (Algorithm 6 line 2). This reduces the filter to a more conventional cross-bilateral filter over the incident radiance, without the need to compute W_p^K. For smooth regions this improves the run time of the filter without loss in quality, since the Poisson weights are unnecessary due to the absence of discontinuities in the illumination.

6.4 Results and discussion

To evaluate the technique proposed here, numerical and qualitative results are presented for scenes containing a wide range of materials and a varying array of noise characteristics. Alongside this, complex texture maps are rendered which typically present difficulties to image-based filtering techniques when combined with noisy illumination. Comparisons with unbiased path tracing and a geometry aware cross-bilateral filter are also presented. Visual, qualitative comparisons show the improvements and reduction of general Monte Carlo noise and spiked noise brought about by the novel filtering techniques.

The same path tracing implementation used in the previous chapter (Section 5.7), is used once again for comparisons with an unbiased technique and as it is the basis of the renderer in the irradiance aware filtering scheme. This includes the variance reduction techniques, making use of next event estimation with multiple importance sampling for direct lighting at each vertex, BRDF importance sampling and Russian roulette for efficient path extension. Images are rendered at 768x768 resolution and tone mapping is applied as before, as described in [RSSF02]. The same hardware is also used for the rendering comprising of a 2.66Ghz Intel Core i7 CPU utilising 8 threads, and 8GB of RAM.

Reference (20,224 spp)

Path Tracing (21 samples) | Our Method (16 samples) | Cross-Bilateral (Smoothing noise) | Cross-Bilateral (Preserve texture)

Figure 6.15: *Glossy spheres test scene, comprised of multiple glossy materials and high detail textures. Path tracing and our method are compared after equal rendering times (see Table 6.1), and the cross-bilateral filter uses the same input samples as the our method*

Unlike the previous chapter, where images were rendered for up to thirty minutes, here rendering times are reduced to the order of one minute. The motivation behind this is that image based filtering serves to provide fast previews with similar appearance to the final rendering. Since the largest bias is introduced early on, in place of additional path tracing computations, short rendering times show off the capabilities of the filtering schemes when provided with a minimal quantity of unbiased input data. Overtime the image will reflect the abilities of the underlying path tracer as opposed to the image based filtering techniques.

To demonstrate the generality of parameter selection, the same settings are used for all images in this chapter. Clustering sensitivity is dictated by μ, set to 0.5. The kernel bandwidth over the pixel neighbourhood is set to $\sigma = 11$ for use in the high intensity noise removal and the probabilistic convolution filtering. The high intensity noise removal procedure uses a low density threshold, $d = \frac{4}{\sigma^2 \mu}$, removing contributions with a local density below this threshold. This has aim of removing only the highest energy samples that could not otherwise be filtered using convolution based approaches, and thus maximising the energy preserved by this filtering stage. For comparisons with a standard cross-bilateral filter, direct lighting is not filtered on specular or glossy surfaces, and the same kernel size, texture, depth and normal buffers are constructed for both methods. The layered texture buffer used in the proposed illumination aware filtering technique is collapsed into a single texture buffer for use with the cross-bilateral filter.

Metallic spheres Figure 6.15 shows a scene with highly detailed, realistic textures in addition to diffuse, pure-specular and glossy surfaces. The glossy materials are based on reflectance data from aluminium and copper metals as discussed in [PH10] with varying degrees of roughness. The scene is lit by three small area light sources that provide hard and soft shadow gradients, and regions of multiple overlapping shadows. The variety of reflective materials provide indirect views of the surrounding high detail textures and a large number of glossy inter-reflections are created that are typically challenging for image-based filters.

Rendering this scene with path tracing alone produces high noise levels due to the high throughput of the glossy materials, resulting in speckled noise across the image when combined with BRDF importance sampling from diffuse surfaces. Detail in the multiple reflections is obfuscated by noise, making the edges in reflections difficult to identify.

Results presented for the cross-bilateral filter use two sets of parameters. The first is tweaked post-render in an attempt to visibly reduce noise. This is achieved by maintaining the same kernel bandwidth σ for the image space neighbourhood, and adjusting the range buffer kernel over the texture colour domain to provide less restrictive filtering across pixel intensities in the texture buffer. Thus, more pixels in the kernel contribute to the target pixel in an effort to improve the signal to noise ratio. The second uses a narrow kernel over the texture buffer, to try and preserve maximum detail in the specular reflections and visible diffuse surfaces.

The cross-bilateral filter reduces the high frequency noise, but at the cost of introducing significant filtering artefacts, blurring textures significantly. Shadow boundaries and strong reflections are not represented in the low variance geometry range buffers, and thus cannot be preserved effectively no matter the parameter choice. In the second instance where tighter restrictions on the texture buffer kernel are imposed, the cross-bilateral filter fails to reduce

Scene		Samples	Path Tracing	Noise Removal	Filtering	Total Time
Glossy Spheres	*PT*	22	44.2s	-	-	44.2s
	Our Method	16	32.7s	0.51s	8.6s	41.8s
Ajax in Water	*PT*	21	45.2s	-	-	45.2s
	Our Method	16	35.1s	0.74s	7.4s	43.24s

Table 6.1: *Time spent in each stage of the irradiance aware filtering algorithm with reference to path tracing. The path tracing stage for our method includes management of the sample clusters, and generation of the geometry and texture buffers.*

noise across small texture features, like the mortar in the brickwork and reflection of the panels in the wooden floor, since in the kernel is too restricted by the texture detail to suitably distribute the energy across the image.

Applying the noise removal and filter proposed in this chapter provides a drastic reduction in noise across the image, whilst preserving all types of visible edges. Using the Poisson weights for edge detection avoids overly blurring high frequency features such as the overlapping direct shadows on the floor, producing results that closely match the referenced path tracing image. Ignoring the texture detail during filtering, operating over the incident radiance ensures that the texture detail is preserved and completely unaffected by the filtering process, regardless of parameter settings. This can become especially important for specular reflections, where texture detail becomes finer due to convex reflection and filters operating over radiance samples are forced to reduce the filter bandwidth.

Ajax bust in water The second scene (Figure 6.16) provides a variety of effects difficult for both filtering algorithms and path tracing. The well known Ajax bust (donated to the community by Jotero.com [Jot13]) provides complex and intricate geometry lit from above by a large area light source. This creates self shadowing, allowing the detail in the geometry to be seen and the peak of the hat shades the left side of the face. The lower portion of the scene presents difficult conditions for path tracing due to the specular reflection and refraction on the surface of the water. In addition, the marble checkerboard floor presents further challenges for radiance and image based filtering techniques as it contains subtle detail, especially problematic viewed via refraction on the water surface.

Path tracing is again riddled with noise from the presence of specular BRDFs. Being shaded by the peak of the hat, geometry in the face is reliant on indirect illumination from the environment. Containing noticeable levels of noise, the geometry detail is masked and difficult to distinguish. Relying on implicit radiance contributions from the specular interactions, path tracing struggles to evaluate samples on and under the surface of the water, with high levels of noise making some regions indistinguishable.

Using the novel illumination aware filtering scheme minimises the blurring of self-shadows on the face, providing more depth and clarity to the beard and hair especially in regions lit indirectly. Refraction through the specular surface of the water is noticeably improved, with fewer artefacts even with the sparsity of the samples where the path tracing input is plagued

Reference (20,224 samples)

Reference | Path Tracing (21 spp) | Our Method (16 spp) | Cross-Bilateral

Figure 6.16: *Ajax bust in water, displaying complex geometry, transmission, caustics and self shadowing. The images for path tracing and our method are equal time comparisons (see Table 6.1). The cross-bilateral filter uses the same input samples as our method.*

Scene	Total Clusters	Single	Memory	Outliers
Spheres	6.19M	1.67M	236MB	0.109%
Ajax	5.56M	1.91M	223MB	0.274%

Table 6.2: *Clusters and memory statistics for our results. Single denotes the number of clusters representing just one sample, and the last column is the percentage of outlying samples. Memory use includes our texture, depth and normal buffers (29MB per image).*

with noise, and the texture detail barely visible. Filtering incident radiance without restrictions imposed by texture detail ensures that the use of the input samples can be maximised in geometrically similar regions. This preserves texture detail even through the water surface, and the sparse approximation of the illumination is handled better than the cross-bilateral filter, which introduces streaking artefacts along boundaries of the filter kernels.

Chromatic lighting The illumination aware filtering framework can also successfully handle chromatic lighting effects, despite being based on a 1D luminance domain. Figure 6.17 shows renderings of a scene with multiple light sources of varying intensities, colours and sizes. The angled grid structure results in numerous overlapping shadows with graduated hard to soft boundaries as the distance between the occluder and the receiver increases.

Since the cross-bilateral filter is ignorant of all illumination boundaries, it blurs across the image and the lack of texture detail does not restrict the kernel, which can in some cases can be useful and reduce the bias introduced over such unidentified edges. Using Poisson weights and the luminance based contribution clustering, the new scheme can preserve illumination edges, differentiating between contributions from each light source and effectively de-noise the image without merging important features.

Memory usage and performance Performance data for the main test scenes are provided in Table 6.1 that presents the breakdown of the rendering time spent in each stage of the process. As the aim is to operate in a progressive rendering paradigm, the vast majority of the computation is taken up by the sample generation and evaluation using path tracing, since this dictates the overall convergence characteristics of the image. The overheads of processing the sample contributions and clusters in the illumination aware filtering is negligible, due to the memory efficient and ordered storage of the clusters, ensuring both insertion and cluster merging is fast. In addition, the high intensity noise removal step also provides minimal overhead due to the optimisations discussed in Section 6.3.1 and the need for only incremental updates at each frame. The remainder of the runtime (around 20-25%) is taken up by the illumination aware convolution filter, dominated by the formation of the kernel groups from existing pixel clusters from which the Poisson weights are derived.

Clustering and memory usage statistics are outlined in Table 6.2. The increased complexity of the glossy spheres scene, as a result of overlapping shadows and complex reflections, requires additional clusters in order to preserve detail in the illumination. This is mirrored in the slightly reduced filtering performance, due to the increase in clustering operations. For the

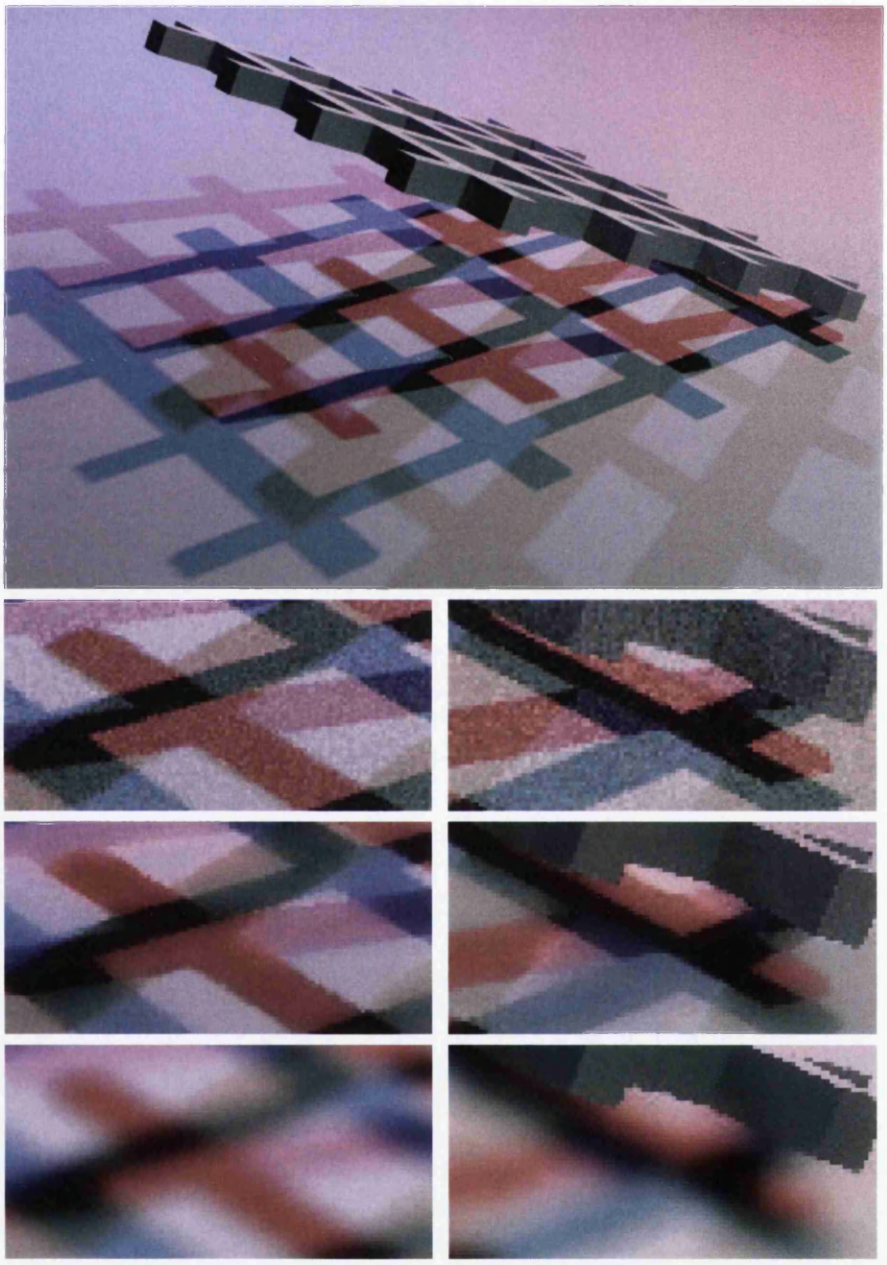


Figure 6.17: *A green grid above a white plane lit by three coloured lights of varying size. Top: Path traced reference. Second row: Path traced input with 16spp. Third row: Our approach, handling the overlapping sources, shadow gradients, and colours. Bottom row: cross-bilateral filter.*

Ajax scene, filtering performance is improved due to the low occlusion in much of the scene, resulting in more uniform lighting which produces fewer changes in the integrand allowing larger clusters across many pixels.

The number of single contribution clusters in the framework follows the level of spiked noise in the image, as each single cluster may represent potential outliers in the sample distribution. Due to the complex specular paths in the Ajax bust scene, the amount of work required by the high intensity noise filter is increased. However it also has a positive impact on cleaning up the overall image, and quantitatively results in a larger number of contributions being classified as outliers and removed from the image.

The net result of is that the memory usage and performance for both scenes is approximately equal, despite the varying illumination conditions and complexity of the image.

6.5 Limitations and future work

It has been shown that the ability to isolate features in the illumination and pixel integrand can be of significant benefit when reducing noise in Monte Carlo estimators. The clustering framework proposed here can deal with complex light transport, texture detail and geometry while providing high quality results. There are however some limitations that are unaddressed in the current work and require further investigation to extend the success of the techniques proposed here.

In the current implementation outlined here there are a number of opportunities for performance improvements. As the discontinuities contained within the clustering framework aim to preserve only visible edges in the image, it seems logical to turn towards perceptually based metrics to improve the resulting clusters. Given the current sample distribution across a pixel for an iteration of the algorithm, a perceptual model could be employed, for example derived from the ideas established by Ferwerda et al.[FPSG96], that controls the merging of clusters based on their perceptual difference, as opposed to the heuristics proposed here. Such approaches have been successfully applied in the past to the adaptive placement of samples over the image plane (for example by Ramasubramanian et al.[RPG99]), using perceptual analysis of the image. To this end, a more optimal cluster set could be found that could in turn reduce the computation required during all stages of the framework, providing faster filtering performance.

Parameter and kernel bandwidth selection Although it has been shown that good results are achieved with constant parameter selection, fully automatic methods are desirable especially for progressive rendering. Moving to perceptual metrics could reduce the need for an explicit sensitivity threshold μ, but the kernel bandwidth σ and density threshold d parameters remain.

Optimal bandwidth selection is still an open problem, as discussed in relation to photon density estimation in Chapter 4. In the context of illumination-aware filtering, addressing this problem of finding optimal per-pixel and layer specific kernel bandwidths could provide improvements in the balance between noise, bias and performance.

Currently, a single bandwidth is used globally across all layers in all pixels. As a result the potential for performance gains is large for regions that have little noise, since the bandwidth can be reduced minimising the number of pixels that need processing without visible differences. Conversely for regions with high levels of noise, the bandwidth may need to be increased to reduce variance, allowing improved image quality from the same overall computation. The introduction of a layer based approach means that these potential gains are also increased. For pixels in direct shadow (such as the ceilings in the glossy spheres scene, and the ceiling and floor of the Ajax scene which is underwater) no filtering is necessary for the direct lighting layer since there are no contributions from pixels in the kernel bandwidth.

One intuitive approach to bandwidth selection is to iteratively increase the kernel bandwidth during each filter pass, up to some stopping criteria. Classical photon mapping employs such a strategy to ensure enough photons are included in the radiance estimate to avoid introducing variance. For image based filtering however, the goal is to minimise noise in the target pixel, which is reliant on the filtering weights chosen for individual pixels in the kernel. Estimating the weights across the kernel prior to fixing the bandwidth is difficult to do efficiently, especially for complex cases where the final weights are a result of multiple terms, such as the geometry data and Poisson calculations required by this work. An alternative approach is to adjust the kernel bandwidth based on the estimated variance or error of the previous filter iteration. This introduces new problems, such as accounting for the change in variance brought about by the newly generated samples, which an estimated kernel bandwidth may not account for. Automatic bandwidth selection would also provide an automated means for convergence, currently achieved by manual reduction of σ over the image plane, which is not ideal.

Aliasing and geometry Operating over the incident radiance of vertex contributions allows filtering to preserve the texture detail of a pixel. However, relying on a cross-bilateral filter also requires depth and normal buffers to determine discontinuities in the geometry. As a result, geometric aliasing is apparent (see the edges of the spheres in Figure 6.15) due to the restriction of the filter kernel. This is a problem common to many image based filtering methods, and the approach presented here is no different. Recent works such as anti-aliasing recovery by Yang et al.[YSLH11] have been adopted to counteract this problem of geometric aliasing, and has been successfully applied to the cross-bilateral filter by Schwenk et al.[SBF13]. Although removing restrictions over texture range buffers in this work, the principle problem of restriction in other domains still remains, reducing efficiency across areas with fine geometry.

As the clustering framework and filtering steps are somewhat independent, it would be interesting to use alternative image based techniques aside from the cross-bilateral filter. Buades non-local means filter [BCM05b] may further improve this layered approach without the restrictions and limitations imposed by the cross-bilateral filter.

Extensions As with any filtering approach, the quality of the output is reliant on the input samples provided by the renderer. It is therefore useful to construct a filtering scheme that is orthogonal to existing variance and noise reduction techniques. Adaptive rendering, based on adjusting the sample distribution over the image plane, could be implemented orthogonally to irradiance-aware filtering, with minor changes. Using the Poisson distribution provides the

likelihood of an event based on the mean occurrence of events within a given interval. For uniform sampling, the interval (ie: number of rendered samples) in each pixel is equal, hence comparison is simple. Under an adaptive scheme, each interval is dependant on the adaptive distribution. Thus, the probabilities of each pixel can be normalised with respect to the pixels sample density before computing Poisson weights.

With regards to mutation based adaptive sampling such as Metropolis light transport, samples are not distributed according to constant PDFs, since the PDF is build over the course of rendering (discussed in Chapter 4). Thus application of a probabilistic filtering scheme is non-obvious in such a context, and is yet to be explored.

Similarly, investigation into the use of depth of field and motion blur effects has not been established. Since the samples are clustered based on luminance, such effects should fit into the existing framework with respect to preservation of illumination edges. An alternative approach may however be necessary for detecting edges in the geometry, since the range buffers used in the cross-bilateral filter will have increased variance. Recent work based on the cross-bilateral filter by Li et al.[LWC12] and Sen and Darabi [SD12] incorporate these effects by comparing the variance and mean of the sample depths or surface normal values for each pixel. Such an approach could also be applied here, at the cost of small memory overhead.

An obvious approach for further work is to utilise the power of the GPU to investigate possible improvements in performance and scalability of the algorithm. Image based filtering methods are generally well suited to highly parallel architectures due to the independence of each pixel, and predictable computational requirements. Both the clustering and filtering algorithms presented here operate on a per pixel basis, thus the coupling between tasks for each pixel is relatively loose, which should aid scalability.

6.6 Conclusions

Obtaining fast previews of global illumination for virtual environments is a useful property, maximising the amount of feedback about the final measurements or appearance of a scene. However this is only true if such feedback is reliable, that is, it represents the end solution with reasonable accuracy, minimising error and bias. This chapter has presented a clustering and filtering framework that takes a novel multi-layer approach to filtering to achieve such a goal. By effectively utilising intermediate radiance measurements and surface details that are already a by-product of unbiased Monte Carlo rendering, the work presented here has proposed a new approach to enable high quality illumination previews.

The core concepts of this chapter have described a set of techniques that together can effectively reduce noise in Monte Carlo methods under complex lighting conditions whilst preserving image discontinuities from overlapping phenomena in the illumination. Special attention is paid with respect to geometric features, and fine texture detail is fully preserved and does not restrict the abilities of the image based filtering scheme. Thus a number of contributions have been discussed including:

- The introduction of a layer-based clustering framework that accurately represents changes in the illumination in multiple dimensions, both across the image plane and within image pixels themselves.

- A novel probabilistic filtering technique was introduced that preserves the discontinuities in the illumination across the image, maintaining sharp details such as hard shadows and reflections which is a significant challenge in image based filtering.
- Permitted by the clustering framework, the filtering scheme operates independently across each illumination layer, improving the ability to filter complex overlapping effects without reducing the filter kernel's effectiveness.
- A density based noise-removal technique for highly efficient removal of statistical outliers in the sample data, eliminating high intensity noise common in Monte Carlo renderers.
- Finally, both filtering and noise-removal can be fine tuned post-render making it especially effective in the trade off between variance and bias, depending on user requirements.

The clustering framework is the core concept, isolating samples according to particular characteristics (in this case path length and luminosity) and separating the distinctive but overlapping features of the integral. On top of this, a high intensity noise removal procedure was introduced in order to remove spiked noise caused by statistical outliers, tailored towards a fast pre-process for image based convolution. It is believed that both of these techniques can be applied in other disciplines, in order to probabilistically separate set of samples from complex integrals evaluated using stochastic methods. Thus, such a clustering procedure can enable further techniques, aside from the cross-bilateral filter discussed here, to remove noise and improve integrator estimates. The dynamic nature of the clustering procedure enables large quantities of samples to be handled efficiently whilst maintaining discontinuities in regions of importance, and compressing the sample data in those that are not.

Applied to Monte Carlo global illumination methods, the result is a set of techniques that together can effectively reduce noise under complex lighting conditions whilst preserving image discontinuities in addition to respecting geometric features. Fine texture detail is fully preserved and does not restrict the abilities of the filtering kernel, thus keeping the size requirements for filtering kernels small.

Chapter 7

Conclusion

Monte Carlo techniques have been at the forefront of photo-realistic global illumination for many decades. Unbiased path tracing techniques have long been known for their high accuracy, robustness and ease of use. Their popularity has been furthered by their strong mathematical foundations and the adoption of techniques from existing disciplines that outlive the use of Monte Carlo techniques in computer graphics.

The renewed attention of researchers in recent years can be attributed in part to the development and widespread availability of parallel architectures; an excellent platform for the highly scalable and progressive nature of path tracing. However, as global illumination has advanced, so have the demands placed upon such algorithms, requiring new approaches for solving light transport in ever more realistic environments.

This thesis has explored several directions of promising research that built upon unbiased Monte Carlo techniques to go some way in addressing the shortfall of research to meet such demands. The novel insights and the resulting algorithms in this thesis have begun to show the power of utilising the wealth of additional information available in path based light transport algorithms. The effectiveness noise reduction brought about by such techniques has been backed-up by both visual and quantitative evidence.

In Chapter 5, the high complexity of light transport in virtual environments was tackled by the careful combination of unbiased path based and progressive density estimation based algorithms. Building upon the natural strengths of each paradigm, the characteristic speckled noise associated with pure Monte Carlo methods was eliminated, whilst the effects of bias typically incurred by density estimation techniques were minimised. This was made possible by utilising the characteristics of each path vertex contribution. Modifying the SPPM algorithm to handle the caustic path-space allows the computation of specular-diffuse-specular lighting that is otherwise difficult for unbiased vertex based methods. Regions of sparse photon distributions problematic during density estimation occur in the presence of low frequency BRDFs, which are effectively handled by the explicit connections of next event estimation in path tracing. The advantages of such an approach were demonstrated with a number of test scenes, providing compelling visual and quantitative comparisons. Furthermore, the progressive nature of both techniques is maintained, allowing both visually pleasing illumination previews and high quality final results.

Extracting and analysing the vertex data of radiance contributions during the course of this work allowed several insights to be made into the distribution of features and noise within the space of the rendering integral. Treating path space as a series of layers across path lengths provided additional insight into the uses of such path data in a progressive framework. Although producing high quality results for complex illumination, both Monte Carlo techniques and density estimation suffer from significant variance and bias removed only with large computational effort. Additional sampling of the integral eventually reduces noise, but the temporary and controlled introduction of bias in place of variance can provide more pleasing previews at low sample counts. Numerous such image-based filtering techniques exist but many lack the generality or accuracy to be an effective partner to Monte Carlo rendering.

Chapter 6 introduced a novel image-based filtering approach supplemented by path vertex data to build an irradiance aware probabilistic filtering framework. First, an efficient and dynamic clustering framework was proposed in order to compactly store vertex contributions, whilst preserving information about illumination discontinuities in the image. Over this framework, a highly efficient technique for the removal of spiked noise was proposed, acting as both a pre-process to prevent filtering artefacts, and a probabilistic convolution filter that preserves a wide range of features in the estimators of unbiased renderers. The culmination of these novel contributions lead to a reduction in noise levels for a number of scenes, while preserving high frequency texture, geometry and illumination features which can otherwise be challenging to identify using convolution filters.

Open problems and future research

The new concepts developed in this thesis have provided a initial step into understanding the importance of leveraging path vertex data to reduce variance, eliminate noise and improve visual artefacts prevalent in Monte Carlo global illumination techniques. Novel approaches to both high quality progressive rendering and effective image based filters have been presented which successfully utilise path vertex data readily available during rendering.

Glossy materials present a problem for many global illumination techniques, including to an extent the work presented here. For the hybrid approach in Chapter 5, the handling of such materials is non-optimal. When scattering from glossy materials with narrow BRDFs, density estimation performs poorly due to its heavy reliance on the incident direction of photons. Such BRDFs also restrict the explicit camera connections of next event estimation in unbiased techniques, causing greater challenges under more difficult lighting conditions.

The more general problem of caustic illumination visible via glossy scattering is a long and outstanding one, but can be improved by the use of recent Markov-chain MC techniques [JM12]. Subsequent to the publication of work in this thesis (in [DJM12]), researchers have further addressed the problems of combining unbiased Monte Carlo and density estimation techniques from a theoretical standpoint, reinforcing the evidence that the combination of path based and density estimation based approaches is an important topic for future research [HPJ12, GKDS12].

In addition, research built upon these two predominant paradigms still faces limitations inherent to each approach. Stochastic based sampling is still relied upon to locate important regions of the integrand, which can only be effectively exploited by importance based and

Markov-chain MC methods once found. Initial bias and kernel bandwidth selection for density estimation is still an open problem despite seeing renewed attention in light of the popularity of progressive photon mapping and its extensions. Despite being desirable for low variance previews, the initial bias of density estimation approaches is persistent throughout the rendering, and thus research into more effective ways of minimising such bias without introducing variance is desirable. Utilising path vertex data to separate illumination features in the gather kernel may in turn allow the detection of estimates with high proximity or boundary bias which can be segregated from the remaining low bias estimates.

Similarly, the poor efficiency of progressive density estimation as the bandwidth radius is reduced is yet to be tackled. Even for techniques that combine multiple paradigms, the convergence of such approaches is still dominated by the poorest of the estimators [HPJ12].

Layer based separation of path contributions has been shown to separate illumination features of the integrand while maximising the use of contributions within an image based neighbourhood. Having lifted the restrictions imposed by relying on a texture based range buffer, it would be desirable to eliminate similar restrictions imposed by geometry buffers based on depth and normal data. An extension into world space is a potential research direction which could extend the neighbourhood from 2D image space to identify features in world space, providing convolution filtering over spatially distant pixels across the image, similar in principle to non-local means.

Investigation into the application of probabilistic and layer based techniques for photon mapping is also yet to be explored. The use of Monte Carlo techniques for unbiased generation of photons means that probabilistic methods are naturally suited to photon mapping and virtual point light methods, in addition to the camera based path tracing that has been the focus of this work. Photon mapping could also benefit from statistically based techniques to reduce error or visual artefacts.

A vast amount of additional vertex data is available during Monte Carlo rendering that is yet to be properly investigated. For example, incident radiance directions can also play a part in the identification and isolation of illumination features, a critical part of low-bias noise reduction especially for glossy materials. Such data is yet to be leveraged for use in progressive rendering. Developing more advanced visualisation practical tools to aid general investigation and analysis of empirical path space in a more general way could be of great benefit to researchers. Development of a flexible sample storage framework coupled with more advanced data visualisation techniques could aid a more exploratory approach to light transport research. Such tools have been helpful in identifying path correlation and empirical observation of path vertex data that has proven useful throughout the development of work in this thesis. Extending these tools may aid the development of future research along similar themes.

In a wider scope, the wealth of existing ideas need to make their way into widespread use through the combination of orthogonal techniques in efficient and complimentary ways. Though a wide range of new research techniques are presented annually, many are seldom seen organised under a single framework. Drawing from avenues such as:

- Image based filtering,
- Image based adaptive rendering,

- Object space importance caching,
- Markov-chain Monte Carlo integration

can together provide benefits to one another. The work presented here in both Chapter 5 and Chapter 6 are theoretically orthogonal one another, in addition to adaptive rendering and importance caching approaches. Combining multiple techniques efficiently can focus on regions of the integrand with high importance and variance, but also target more specific layers or features of illumination instead of relying on variance at a pixel level. Additionally, identifying regions in which low cost solutions (such as image based filtering techniques) are perceptually viable can improve noise efficiently, allowing the more computationally expensive techniques such as adaptive sampling to be concentrated in regions where less costly techniques are impractical or would produce significant artefacts. Thus, not only is there room for investigation into how such techniques can benefit one another, but also into automatic methods for identifying under what conditions do they perform poorly in comparison to one another such that their strengths can be combined more efficiently.

Even throughout the period of research culminating in this thesis, the development of work aimed towards solving and understanding global illumination both in academia and industry has made tremendous progress. It is clear from the discussion and development of ideas in this thesis that despite decades of research there are still a great many problems in photo-realistic rendering that remain open. From radiometric effects and natural phenomena (Chapter 1) yet to see widespread use in rendering, to Monte Carlo variance reduction techniques established in external disciplines and mathematics (Chapter 3) that may have potential benefits for computer graphics applications. There are a great many avenues of research that can close the gap between photographic images and physically-based rendering, and with new research emerges new questions, as this thesis has shown.

Global illumination research is driven by the demand for realism and speed. In making steps to fulfil such demands, it is the hope that this thesis can help further the application and popularity of physically-based, highly accurate global illumination. Increasing the awareness and elegance of physically-based rendering is also an important target for research, pushing the funding and support for the development of new algorithms and purpose built hardware that will help ray tracing techniques achieve the leaps and bounds that rasterisation has gone through over the last decades. In turn, the constant evolution of hardware provokes new research and ideas that are constantly shifting the boundaries and possibilities for what is an ever more complex and far reaching topic. The research described in this thesis has gone some way to explore and expand just a few of the many possible avenues that will inevitably lead to photo-realistic, physically-based global illumination techniques becoming the cornerstone of rendering across a wide number of industries.

Bibliography

[AGDL09] Andrew Adams, Natasha Gelfand, Jennifer Dolson, and Marc Levoy. Gaussian KD-Trees for Fast High-Dimensional Filtering. *ACM Trans. Graph.*, 28(3):1–12, 2009. 15

[AK90] James Arvo and David Kirk. Particle transport and image synthesis. *SIGGRAPH Comput. Graph.*, 24(4):63–66, September 1990. 37, 42

[APSS01] Michael Ashikhmin, Simon Premoze, Peter Shirley, and Brian E. Smits. A variance analysis of the metropolis light transport algorithm. *Computers & Graphics*, 25(2):287–294, 2001. 58

[AS00] Michael Ashikhmin and Peter Shirley. An anisotropic phong brdf model. *Journal Graph. Tools*, 5(2):25–32, February 2000. 20

[BAJ08] B. C. Budge, John C. Anderson, and Kenneth I. Joy. Caustic forecasting: Unbiased estimation of caustic lighting for global illumination. *Comput. Graph. Forum*, 27(7):1963–1970, 2008. 74, 114

[BB07] David C Banks and Kevin Beason. Fast global illumination for visualizing isosurfaces with a 3d illumination grid. *Computing in Science & Engineering*, 9(1):48–54, 2007. 2

[BCM05a] A. Buades, B. Coll, and J. M. Morel. A Review of Image Denoising Algorithms, with a New One. *Multiscale Modeling & Simulation*, 4(2):490–530, 2005. 78

[BCM05b] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. A non-local algorithm for image denoising. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2 of *CVPR '05*, pages 60–65, Washington, DC, USA, 2005. IEEE Computer Society. 81, 157

[BEL+07] Solomon Boulos, Dave Edwards, J. Dylan Lacewell, Joe Kniss, Jan Kautz, Peter Shirley, and Ingo Wald. Packet-Based Whitted and Distribution Ray tracing. In *Proc. Graphics Interface*, pages 177–184, New York, NY, USA, 2007. ACM. 17

[BEM11] Pablo Bauszat, Martin Eisemann, and Marcus Magnor. Guided image filtering for interactive high-quality global illumination. *Comput. Graph. Forum (Proc. Eurographics Symposium on Rendering)*, 30(4):1361–1368, June 2011. 79, 122

[Ben75] Jon Louis Bentley. Multidimensional Binary Search Trees used for Associative Searching. *Commun. ACM*, 18(9):509–517, 1975. 15

[Bik13] Jakko Bikker. *Real-time Path Tracing in Games*. PhD thesis, Technische Universiteit Delft, 2013. 2

[Bli77] James F. Blinn. Models of light reflection for computer synthesized pictures. *SIGGRAPH Comput. Graph.*, 11(2):192–198, July 1977. 18, 20

[BM98] Mark R. Bolin and Gary W. Meyer. A perceptually based adaptive sampling algorithm. In *Proc. Computer graphics and interactive techniques*, SIGGRAPH '98, pages 299–309, New York, NY, USA, 1998. ACM. 76

[BN76] James F. Blinn and Martin E. Newell. Texture and reflection in computer generated images. *Commun. ACM*, 19(10):542–547, October 1976. 18

[BRDC12] Thomas Bashford-Rogers, Kurt Debattista, and Alan Chalmers. A significance cache for accelerating global illumination. *Comp. Graph. Forum*, 31(6):1837–1851, September 2012. 75, 113

[CAE08] David Cline, Daniel Adams, and Parris K. Egbert. Table-driven Adaptive Importance Sampling. *Comput. Graph. Forum*, 27(4):1115–1123, 2008. 75

[CCM03] Alan Chalmers, Kirsten Cater, and David Maflioli. Visual attention models for producing high fidelity graphics efficiently. In *Proc. Spring conference on Computer graphics*, SCCG '03, pages 39–45, New York, NY, USA, 2003. ACM. 12

[CG85] Michael F. Cohen and Donald P. Greenberg. The hemi-cube: a radiosity solution for complex environments. In *Proc. Computer graphics and interactive techniques*, SIGGRAPH '85, pages 31–40, New York, NY, USA, 1985. ACM. 42

[CHB+12] Tom Cuypers, Tom Haber, Philippe Bekaert, Se Baek Oh, and Ramesh Raskar. Reflectance model for diffraction. *ACM Trans. Graph.*, 31(5):122:1–122:11, September 2012. 8

[Che81] R. C.H. Cheng. The use of antithetic control variates in computer simulations. In *Proc. Conference on Winter simulation*, volume 1, pages 313–318, Piscataway, NJ, USA, 1981. IEEE. 40

[CPC84] Robert L. Cook, Thomas Porter, and Loren Carpenter. Distributed Ray Tracing. *SIGGRAPH Comput. Graph.*, 18(3):137–145, 1984. 24, 42

[CRMT91] Shenchang Eric Chen, Holly E. Rushmeier, Gavin Miller, and Douglass Turner. A progressive multi-pass method for global illumination. In *Proc. Computer graphics and interactive techniques*, SIGGRAPH '91, pages 165–174, New York, NY, USA, 1991. ACM. 90

[Cry13] Crytek GmbH. Cryengine visuals. `http://www.mycryengine.com/index.php?conid=8`, February 2013. 3

[CTE05] David Cline, Justin Talbot, and Parris Egbert. Energy redistribution path tracing. *ACM Trans. Graph.*, 24(3):1186–1195, July 2005. 59, 74

[DGR+09] Zhao Dong, Thorsten Grosch, Tobias Ritschel, Jan Kautz, and Hans-Peter Seidel. Real-time indirect illumination with clustered visibility. In *Vision, Modeling, and Visualization Workshop*, 2009. 73

[DHK08] Holger Dammertz, Johannes Hanika, and Alexander Keller. Shallow Bounding Volume Hierarchies for Fast SIMD Ray Tracing of Incoherent Rays. *Comput. Graph. Forum*, 27(4):1225, 2008. 17

[Dig12] Digital Cinema Initiatives, LLC. DCI Specification, Version 1.2 with Errata as of 30 August 2012 incorporated. http://www.dcimovies.com/specification/, August 2012. 2

[DJ13] Ian C. Doidge and Mark W. Jones. Probabilistic illumination-aware filtering for monte carlo rendering. *The Visual Computer*, 29(6-8):707–716, 2013. 123

[DJM12] Ian C. Doidge, Mark W. Jones, and Benjamin Mora. Mixing monte carlo and progressive rendering for improved global illumination. *The Visual Computer*, 28(6-8):603–612, 2012. 85, 112, 162

[DK08] Holger Dammertz and Alexander Keller. The Edge Volume Heuristic - Robust Triangle Subdivision for Improved BVH Performance. In *IEEE Symposium on Interactive Ray Tracing*, page 155, 2008. 16

[DKH09] Peter Djeu, Sean Keely, and Warren Hunt. Accelerating Shadow Rays Using Volumetric Occluders and Modified kD-Tree Traversal. In *Proc. Conference on High Performance Graphics*, pages 69–76, New York, NY, USA, 2009. ACM. 17

[DKL10] H. Dammertz, A. Keller, and H. P. A. Lensch. Progressive point-light-based global illumination. *Comput. Graph. Forum*, 29(8):2504–2515, 2010. 91, 112, 114

[DLW93] Philip Dutré, Eric P. Lafortune, and Yves Willems. Monte Carlo light tracing with direct computation of pixel intensities. In *International Conference on Computational Graphics and Visualisation Techniques*, pages 128–137, December 1993. 52

[DSHL10] Holger Dammertz, Daniel Sewtz, Johannes Hanika, and Hendrik Lensch. Edge-avoiding a-trous wavelet transform for fast global illumination filtering. In *Proc. High Performance Graphics*, pages 67–75, 2010. 80

[DWB+06] Michael Donikian, Bruce Walter, Kavita Bala, Sebastian Fernandez, and Donald P. Greenberg. Accurate direct illumination using iterative adaptive sampling. *IEEE Trans. on Vis. and Computer Graphics*, 12:353–364, May 2006. 114

[DWR10] Christopher DeCoro, Tim Weyrich, and Szymon Rusinkiewicz. Density-based outlier rejection in monte carlo rendering. *Comput. Graph. Forum (Proc. Pacific Graphics)*, 29(7):2119–2125, September 2010. 78, 133, 136

[ED04] Elmar Eisemann and Frédo Durand. Flash photography enhancement via intrinsic relighting. *ACM Trans. Graph.*, 23(3):673–678, August 2004. 79, 80

[EG07] Manfred Ernst and Gunther Greiner. Early Split Clipping for Bounding Volume Hierarchies. In *IEEE Symposium on Interactive Ray Tracing*, page 73, 2007. 16

[FBF77] Jerome H. Friedman, Jon Louis Bentley, and Raphael Ari Finkel. An Algorithm for Finding Best Matches in Logarithmic Expected Time. *ACM Trans. Math. Softw.*, 3(3):209–226, 1977. 15

[FCH+06] Shaohua Fan, Stephen Chenney, Bo Hu, Kam-Wah Tsui, and Yu-Chi Lai. Optimizing control variate estimators for rendering. *Comput. Graph. Forum*, 25(3):351–358, 2006. 40

[FCL05] Shaohua Fan, Stephen Chenney, and Yu Chi Lai. Metropolis Photon Sampling with Optional User Guidance. In *Proc. Eurographics Symposium on Rendering*, pages 127–138. Eurographics Association, 2005. 66

[FPSG96] James A. Ferwerda, Sumanta N. Pattanaik, Peter Shirley, and Donald P. Greenberg. A model of visual adaptation for realistic image synthesis. In *Proc. Computer graphics and interactive techniques*, SIGGRAPH '96, pages 249–258, New York, NY, USA, 1996. ACM. 12, 156

[FS88] Donald S. Fussell and K. R. Subramanian. Fast Ray Tracing Using K-d Trees. Technical report, University of Texas at Austin, Austin, TX, USA, 1988. 16

[Gar08] Kirill Garanzha. Efficient Clustered BVH Update Algorithm for Highly-Dynamic Models. In *Proc. IEEE Symposium on Interactive Ray Tracing*, 2008. 16

[GBP08] Christiaan P. Gribble, Carson Brownlee, and Steven G. Parker. Practical global illumination for interactive particle visualization. *Computers & Graphics*, 32(1):14–24, 2008. 2

[GKB09] Václav Gassenbauer, Jaroslav Křivánek, and Kadi Bouatouch. Spatial directional radiance caching. *Comput Graph. Forum*, 28(4):1189–1198, 2009. 77

[GKDS12] Iliyan Georgiev, Jaroslav Křivánek, Tomáš Davidovič, and Philipp Slusallek. Light transport simulation with vertex connection and merging. *ACM Trans. Graph.*, 31(6):192:1–192:10, November 2012. 112, 113, 162

[GKPS12] Iliyan Georgiev, Jaroslav Kivnek, Stefan Popov, and Philipp Slusallek. Importance caching for complex illumination. *Comput. Graph. Forum*, 31(2pt3):701–710, 2012. 75

[GL10] Kirill Garanzha and Charles Loop. Fast Ray Sorting and Breadth-First Packet Traversal for GPU Ray Tracing. In *Eurographics*, May 2010. 17

[Gla95] A.S. Glassner. *Principles of digital image synthesis: Vol. 1.* Morgan Kaufmann series in data management systems. Morgan Kaufmann Publishers, 1995. 7, 8, 129

[Gor97] Hugh Gordon. *Discrete Probability.* Springer-Verlag New York, Inc., 1997. 124

[GPSS07] Johannes Gunther, Stefan Popov, Hans-Peter Seidel, and Philipp Slusallek. Realtime Ray Tracing on GPU with BVH-based Packet Traversal. In *Proc. Symposium on Interactive Ray Tracing*, pages 113–118, Washington, DC, USA, 2007. IEEE Computer Society. 17

[GS87] Jeffrey Goldsmith and John Salmon. Automatic Creation of Object Hierarchies for Ray Tracing. *IEEE Comput. Graph. Appl.*, 7(5):14–20, 1987. 16

[GTGB84] Cindy M. Goral, Kenneth E. Torrance, Donald P. Greenberg, and Bennett Battaile. Modeling the interaction of light between diffuse surfaces. In *Proc. Computer graphics and interactive techniques*, SIGGRAPH '84, pages 213–222, New York, NY, USA, 1984. ACM. 42

[Hav00] Vlastimil Havran. *Heuristic Ray Shooting Algorithms.* Ph.d. thesis, Department of Computer Science and Engineering, Faculty of Electrical Engineering, Czech Technical University in Prague, November 2000. 15, 16

[Hec90] Paul S. Heckbert. Adaptive radiosity textures for bidirectional ray tracing. In *Proc. Computer graphics and interactive techniques*, SIGGRAPH '90, pages 145–154, New York, NY, USA, 1990. ACM. 23, 93

[Hes97] T. Hesterberg. Weighted average importance sampling and defensive mixture distributions. *Technometrics*, 37(2):185–194, 1997. 35

[HESL11] Matthias B. Hullin, Elmar Eisemann, Hans-Peter Seidel, and Sungkil Lee. Physically-based real-time lens flare rendering. *ACM Trans. Graph. (Proc. SIGGRAPH 2011)*, 30(4):108:1–108:9, 2011. 10

[HH10] Jared Hoberock and John C. Hart. Arbitrary importance functions for metropolis light transport. *Comput. Graph. Forum*, 29(6):1993–2003, 2010. 59, 76

[HHS05] Vlastimil Havran, Robert Herzog, and Hans-Peter Seidel. Fast Final Gathering via Reverse Photon Mapping. *Computer Graphics Forum (Proceedings of Eurographics 2005)*, 24(3):323–333, August 2005. 66, 67

[HJ09] Toshiya Hachisuka and Henrik Wann Jensen. Stochastic progressive photon mapping. *ACM Trans. Graph.*, 28(5):141:1–141:8, December 2009. 69, 70, 85, 102

[HJ11] Toshiya Hachisuka and Henrik Wann Jensen. Robust adaptive photon tracing using photon path visibility. *ACM Trans. Graph.*, 30(5):114:1–114:11, October 2011. 72, 113

[HJJ10] Toshiya Hachisuka, Wojciech Jarosz, and Henrik Wann Jensen. A progressive error estimation framework for photon density estimation. *ACM Trans. Graph.*, 29(6):144:1–144:12, December 2010. 72, 113

[HJW+08] Toshiya Hachisuka, Wojciech Jarosz, Richard Peter Weistroffer, Kevin Dale, Greg Humphreys, Matthias Zwicker, and Henrik Wann Jensen. Multidimensional Adaptive Sampling and Reconstruction for Ray Tracing. *ACM Trans. Graph.*, 27(3):1–10, 2008. 76

[HKRS02] Jim Hurley, Alexander Kapustin, Alexander Reshetov, and Alexei Soupikov. Fast Ray Tracing for Modern General Purpose CPU. In *Proc. of Graphicon*, 2002. 16

[HKWB09] Miloš Hašan, Jaroslav Křivánek, Bruce Walter, and Kavita Bala. Virtual spherical lights for many-light rendering of glossy scenes. *ACM Trans. Graph.*, 28(5):143:1–143:6, December 2009. 73

[HM08a] Warren Hunt and William R. Mark. Adaptive acceleration structures in perspective space. In *Symposium on Interactive Ray Tracing*, pages 11–17. IEEE, 2008. 17

[HM08b] Warren Hunt and William R. Mark. Ray-Specialized acceleration structures for ray tracing. In *Symposium on Interactive Ray Tracing*, pages 3–10. IEEE, 2008. 17

[HMS05] Jörg Haber, Marcus Magnor, and Hans-Peter Seidel. Physically-based simulation of twilight phenomena. *ACM Trans. Graph.*, 24(4):1353–1373, October 2005. 7

[HMS06] Warren Hunt, William R. Mark, and Gordon Stoll. Fast kD-tree Construction with an Adaptive Error-Bounded Heuristic. In *Symposium on Interactive Ray Tracing*. IEEE, September 2006. 16

[HOJ08] Toshiya Hachisuka, Shinji Ogaki, and Henrik Wann Jensen. Progressive photon mapping. *ACM Trans. Graph.*, 27(5):130:1–130:8, December 2008. 66, 67, 70, 98

[HP01] Heinrich Hey and Werner Purgathofer. Importance sampling with hemispherical particle footprints. Technical Report TR-186-2-01-05, Institute of Computer Graphics and Algorithms, Vienna University of Technology, January 2001. 74

[HP02] Heinrich Hey and Werner Purgathofer. Advanced radiance estimation for photon map global illumination. *Computer Graphics Forum*, 21(3):66

[HPJ12] Toshiya Hachisuka, Jacopo Pantaleoni, and Henrik Wann Jensen. A path space extension for robust light transport simulation. *ACM Trans. Graph.*, 31(6):191:1–191:10, November 2012. 90, 112, 113, 162, 163

[HS07] Robert Herzog and Hans-Peter Seidel. Lighting details preserving photon density estimation. In *Proc. Pacific Conference on Computer Graphics and Applications*, pages 407–410, Washington, DC, USA, 2007. IEEE Computer Society. 66, 67

[HW12] Lukas Hosek and Alexander Wilkie. An analytic model for full spectral sky-dome radiance. *ACM Trans. Graph.*, 31(4):95:1–95:9, July 2012. 7

[Ige99] Homan Igehy. Tracing ray differentials. In *Proc. Computer graphics and interactive techniques*, SIGGRAPH '99, pages 179–186, New York, NY, USA, 1999. ACM Press/Addison-Wesley Publishing Co. 21

[ISP07] Thiago Ize, Peter Shirley, and Steven Parker. Grid creation strategies for efficient ray tracing. In *Proc. Symposium on Interactive Ray Tracing*, pages 27–32, Washington, DC, USA, 2007. IEEE Computer Society. 17

[Jak13] Wenzel Jakob. Mitsuba physically based renderer. `http://www.mitsuba-renderer.org/`, 2013. [Online; accessed March 2013]. 4

[JC95] Henrik Wann Jensen and Niels Jrgen Christensen. Photon Maps in Bidirectional Monte Carlo Ray Tracing of Complex Objects. *Computers & Graphics*, 19(2):215 – 224, 1995. 62

[JC98] Henrik Wann Jensen and Per H. Christensen. Efficient simulation of light transport in scences with participating media using photon maps. In *Proc. Computer graphics and interactive techniques*, SIGGRAPH '98, pages 311–320, New York, NY, USA, 1998. ACM. 60

[Jen95] Henrik Wann Jensen. Importance driven path tracing using the photon map. In Pat Hanrahan and Werner Purgathofer, editors, *Proc. Eurographics Workshop on rendering techniques*, pages 326–335. Springer, 1995. 74, 113

[Jen96] Henrik Wann Jensen. Global illumination using photon maps. In *Proc. Eurographics Workshop on rendering techniques*, pages 21–30, London, UK, 1996. Springer-Verlag. 60

[Jen01] Henrik Wann Jensen. *Realistic Image Synthesis Using Photon Mapping*. A. K. Peters, Ltd., Natick, MA, USA, 2001. 25, 60, 61, 65, 67, 72, 113

[JKRY12] D. Jonsson, J. Kronander, T. Ropinski, and A. Ynnerman. Historygrams: Enabling interactive global illumination in direct volume rendering using photon mapping. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2364–2371, 2012. 2

[JM12] Wenzel Jakob and Steve Marschner. Manifold exploration: a markov chain monte carlo technique for rendering scenes with difficult specular transport. *ACM Trans. Graph.*, 31(4):58:1–58:13, July 2012. 59, 114, 162

[JMLH01] Henrik Wann Jensen, Stephen R. Marschner, Marc Levoy, and Pat Hanrahan. A practical model for subsurface light transport. In *Proc. Computer graphics and interactive techniques*, SIGGRAPH '01, pages 511–518, New York, NY, USA, 2001. ACM. 7, 60

[JMS96] M. C. Jones, J. S. Marron, and S. J. Sheather. A Brief Survey of Bandwidth Selection for Density Estimation. *Journal of the American Statistical Association*, 91(433):401–407, March 1996. 66

[Jot13] Jotero.com. 3D-Scan and 3D Vermessung. `http://http://forum.jotero.com/`, 2013. [Online; accessed April 2013]. 152

[KA91] David Kirk and James Arvo. Unbiased sampling techniques for image synthesis. *SIGGRAPH Comput. Graph.*, 25(4):153–156, July 1991. 76

[Kaj86] James T. Kajiya. The rendering equation. In *Proc. Computer graphics and interactive techniques*, number 4 in SIGGRAPH '86, pages 143–150, New York, NY, USA, 1986. ACM. 24, 42

[KBS11] Javor Kalojanov, Markus Billeter, and Philipp Slusallek. Two-Level Grids for Ray Tracing on GPUs. In *Proc. Eurographics Conference*, pages 307–314, Llandudno, UK, 2011. Eurographics Association. 17

[KD13] Anton S. Kaplanyan and Carsten Dachsbacher. Adaptive progressive photon mapping. *ACM Trans. Graph.*, 32(2):16:1–16:13, April 2013. 112

[Kel97] Alexander Keller. Instant Radiosity. In *Proc. SIGGRAPH '97, Annual Conference Series*, pages 49–56, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co. 72

[KFY+10] Shota Kanamori, Kazuya Fujiwara, Takahiro Yoshinobu, Bisser Raytchev, Toru Tamaki, and Kazufumi Kaneda. Physically based rendering of rainbows under various atmospheric conditions. In *Proc. Pacific Conference on Computer Graphics and Applications*, Pacific Graphics '10, pages 39–45, Washington, DC, USA, 2010. IEEE Computer Society. 7

[KGPK05] Jaroslav Křivánek, Pascal Gautron, Sumanta Pattanaik, and KadiBouatouch. Radiance caching for efficient global illumination computation. *IEEE Trans. on Visualization and Computer Graphics*, 11(5):550–561, 2005. 77

[KHA+12] Jaroslav Křivánek, Miloš Hašan, Adam Arbree, Carsten Dachsbacher, Alexander Keller, and Bruce Walter. Optimizing realistic rendering with many-light methods. In *ACM SIGGRAPH 2012 Courses*, SIGGRAPH '12, pages 7:1–7:217, New York, NY, USA, 2012. ACM. 73

[Khr13] Khronos Group. OpenCL Overview. `http://www.khronos.org/opencl/`, 2013. [Online; accessed April 2013]. 2

[KK02] Thomas Kollig and Alexander Keller. Efficient multidimensional sampling. *Comput. Graph. Forum*, 21(3):557–563, 2002. 37

[KK04] Thomas Kollig and Alexander Keller. Illumination in the Presence of Weak Singularities. In *Monte Carlo and Quasi-Monte Carlo Methods*, 2004. 73, 75, 114

[KRK04] Janne Kontkanen, Jussi Räsänen, and Alexander Keller. Irradiance filtering for monte carlo ray tracing. In *Monte Carlo and Quasi-Monte Carlo Methods 2004*, pages 259–272. Springer, 2004. 77, 134, 142

[KS97] K. Klimaszewski and T.W. Sederberg. Faster ray tracing using adaptive grids. *Computer Graphics and Applications, IEEE*, 17(1):42–51, 1997. 17

[KS13] Nima Khademi Kalantari and Pradeep Sen. Removing the noise in monte carlo rendering with general image denoising algorithms. *Comput. Graph. Forum*, 32(2pt1):93–102, 2013. 81

[KSKAC02] Csaba Kelemen, Lszl Szirmay-Kalos, Gyrgy Antal, and Ferenc Csonka. A simple and robust mutation strategy for the metropolis light transport algorithm. *Comput. Graph. Forum*, 21(3):531–540, 2002. 57, 58

[KZ11] Claude Knaus and Matthias Zwicker. Progressive photon mapping: A probabilistic approach. *ACM Trans. Graph.*, 30(3):25:1–25:13, May 2011. 72, 90, 111, 112

[LAC+11] Jaakko Lehtinen, Timo Aila, Jiawen Chen, Samuli Laine, and Frédo Durand. Temporal light field reconstruction for rendering distribution effects. *ACM Trans. Graph.*, 30(4), 2011. 76

[LD08] Ares Lagae and Philip Dutré. Compact, fast and robust grids for ray tracing. *Comput. Graph. Forum*, 27(4):1235–1244, June 2008. 17

[LES10] Sungkil Lee, Elmar Eisemann, and Hans-Peter Seidel. Real-Time Lens Blur Effects and Focus Control. *ACM Trans. Graph.*, 29(4):65:1–7, 2010. 10

[LL01] D.K.A. Lynch and W.C. Livingston. *Color And Light In Nature 2nd Edition.* Cambridge University Press, 2001. 7

[Lux13] LuxRender. LuxRender GPL Physically Based Renderer. `http://www.luxrender.net/`, 2013. [Online; accessed March 2013]. 4, 9

[LW93] Eric P. Lafortune and Yves D. Willems. Bi-Directional Path Tracing. In *Proc. Computational Graphics and Vis. Techniques*, pages 145–153, 1993. 54, 56

[LWC12] Tzu-Mao Li, Yu-Ting Wu, and Yung-Yu Chuang. Sure-based optimization for adaptive sampling and reconstruction. *ACM Trans. Graph.*, 31(6):186:1–186:9, November 2012. 76, 77, 80, 158

[MB90] David J. MacDonald and Kellogg S. Booth. Heuristics for Ray Tracing using Space Subdivision. *Vis. Comput.*, 6(3):153–166, 1990. 16

[MFS09] Alex Mendez-Feliu and Mateu Sbert. From obscurances to ambient occlusion: A survey. *The Visual Computer*, 25(2):181–196, 2009. 5

[MHH+12] Stephen McAuley, Stephen Hill, Naty Hoffman, Yoshiharu Gotanda, Brian Smits, Brent Burley, and Adam Martinez. Practical physically based shading in film and game production. In *SIGGRAPH 2012 Courses*, New York, NY, USA, 2012. ACM. 1, 6

[Mit13] Mitsubishi Electric Research Laboratories (MERL). MERL BRDF Database version 2.0. `http://www.merl.com/brdf/`, 2013. 20

[MN88] Don P. Mitchell and Arun N. Netravali. Reconstruction filters in computer-graphics. In *Proc. Computer graphics and interactive techniques*, SIGGRAPH '88, pages 221–228, New York, NY, USA, 1988. ACM. 76

[MPBM03] Wojciech Matusik, Hanspeter Pfister, Matt Brand, and Leonard McMillan. A data-driven reflectance model. *ACM Trans. Graph.*, 22(3):759–769, July 2003. 20

[MTAS01] Karol Myszkowski, Takehiro Tawara, Hiroyuki Akamine, and Hans-Peter Seidel. Perception-guided global illumination solution for animation rendering. In *Proc. Computer graphics and interactive techniques*, SIGGRAPH '01, pages 221–230, New York, NY, USA, 2001. ACM. 12

[NNDJ12a] Jan Novák, Derek Nowrouzezahrai, Carsten Dachsbacher, and Wojciech Jarosz. Progressive virtual beam lights. *Comput. Graph. Forum*, 31(4), June 2012. 73

[NNDJ12b] Jan Novák, Derek Nowrouzezahrai, Carsten Dachsbacher, and Wojciech Jarosz. Virtual ray lights for rendering scenes with participating media. *ACM Trans. Graph.*, 31(4):60:1–60:11, July 2012. 73

[NVI13] NVIDIA Corporation. Nvidia optix ray tracing engine. `https://developer.nvidia.com/optix`, 2013. [Online; accessed March 2013]. 4

[ODR09] Ryan S. Overbeck, Craig Donner, and Ravi Ramamoorthi. Adaptive Wavelet Rendering. *ACM Transactions on Graphics (SIGGRAPH Asia 09)*, 28(5):1–12, 2009. 76

[OKG+10] Se Baek Oh, Sriram Kashyap, Rohit Garg, Sharat Chandran, and Ramesh Raskar. Rendering wave effects with augmented light field. *Comput. Graph. Forum*, 29(2):507–516, 2010. 8

[ON94] Michael Oren and Shree K. Nayar. Generalization of lambert's reflectance model. In *Proc. Computer graphics and interactive techniques*, SIGGRAPH '94, pages 239–246, New York, NY, USA, 1994. ACM. 20

[OZ98] A.B. Owen and Y. Zhou. Safe and effective importance sampling. Technical report, Stanford University, 1998. 35

[PBP11] Anthony Pajot, Loïc Barthe, and Mathias Paulin. Sample-space bright spots removal using density estimation. In *Proc. of Graphics Interface 2011*, pages 159–166, 2011. 78, 133, 136

[PBPP11] Anthony Pajot, Loc Barthe, Mathias Paulin, and Pierre Poulin. Combinatorial bidirectional path-tracing for efficient hybrid cpu-gpu rendering. *Comput. Graph. Forum*, 30(2):315–324, April 2011. 57

[Per85] Ken Perlin. An image synthesizer. *SIGGRAPH Comput. Graph.*, 19(3):287–296, July 1985. 21

[PGSS06] Stefan Popov, Johannes Gunther, Hans-Peter Seidel, and Philipp Slusallek. Experiences with Streaming Construction of SAH KD-Trees. *Symposium on Interactive Ray Tracing*, 0:89–94, September 2006. 16

[PH10] Matt Pharr and Greg Humphreys. *Physically Based Rendering, Second Edition: From Theory To Implementation.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2nd edition, 2010. 37, 151

[PKT09] Sylvain Paris, Pierre Kornprobst, and Jack Tumblin. *Bilateral Filtering: Theory and Applications.* Now Publishers Inc., Hanover, MA, USA, 2009. 79, 80

[PSA+04] Georg Petschnigg, Richard Szeliski, Maneesh Agrawala, Michael Cohen, Hugues Hoppe, and Kentaro Toyama. Digital photography with flash and no-flash image pairs. *ACM Trans. Graph.*, 23(3):664–672, August 2004. 79, 80

[PSS99] A. J. Preetham, Peter Shirley, and Brian Smits. A practical analytic model for daylight. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '99, pages 91–100, New York, NY, USA, 1999. ACM Press/Addison-Wesley Publishing Co. 7

[RDGK12] Tobias Ritschel, Carsten Dachsbacher, Thorsten Grosch, and Jan Kautz. The state of the art in interactive global illumination. *Comput. Graph. Forum*, 31(1):160–188, February 2012. 73

[REK+04] Kirk Riley, David S. Ebert, Martin Kraus, Jerry Tessendorf, and Charles Hansen. Efficient rendering of atmospheric phenomena. In *Proceedings of the Fifteenth Eurographics conference on Rendering Techniques*, EGSR'04, pages 375–386, Aire-la-Ville, Switzerland, Switzerland, 2004. Eurographics Association. 7

[RKZ11] Fabrice Rousselle, Claude Knaus, and Matthias Zwicker. Adaptive sampling and reconstruction using greedy error minimization. *ACM Trans. Graph.*, 30(6):159:1–159:12, December 2011. 76

[RKZ12] Fabrice Rousselle, Claude Knaus, and Matthias Zwicker. Adaptive rendering with non-local means filtering. *ACM Trans. Graph.*, 31(6):195:1–195:11, November 2012. 76, 81

[RPG99] Mahesh Ramasubramanian, Sumanta N. Pattanaik, and Donald P. Greenberg. A perceptually based physical error metric for realistic image synthesis. In *Proc. Computer graphics and interactive techniques*, SIGGRAPH '99, pages 73–82, New York, NY, USA, 1999. ACM Press/Addison-Wesley Publishing Co. 12, 76, 156

[RSH05] Alexander Reshetov, Alexei Soupikov, and Jim Hurley. Multi-level ray tracing algorithm. *ACM Trans. Graph.*, 24(3):1176–1185, July 2005. 17

[RSSF02] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. Photographic tone reproduction for digital images. In *Proc. SIGGRAPH*, pages 267–276. ACM Press, 2002. 101, 149

[SBF13] Karsten Schwenk, Johannes Behr, and DieterW. Fellner. Filtering noise in progressive stochastic ray tracing. *The Visual Computer*, 29(5):359–368, 2013. 157

[Sch03] Roland Schregle. Bias compensation for photon maps. *Comput. Graph. Forum*, 22(4):729–742, 2003. 66

[SD12] Pradeep Sen and Soheil Darabi. On filtering the noise from the random parameters in monte carlo rendering. *ACM Trans. Graph.*, 31(3):18:1–18:15, June 2012. 76, 80, 158

[SF90a] K. R. Subramanian and Donald S. Fussell. A Cost Model for Ray Tracing Hierarchies. Technical report, University of Texas at Austin, Austin, TX, USA, 1990. 16

[SF90b] K. R. Subramanian and Donald S. Fussell. Factors Affecting Performance of Ray Tracing Hierarchies. Technical report, University of Texas at Austin, Austin, TX, USA, 1990. 16

[SFD00] Yinlong Sun, F. David Fracchia, and Mark S. Drew. RENDERING LIGHT DISPERSION WITH A COMPOSITE SPECTRAL MODEL. In *Proceedings of the International Conference on Color in Graphics and Image Processing*, pages 51–56, November 2000. 8

[SFD09] Martin Stich, Heiko Friedrich, and Andreas Dietrich. Spatial Splits in Bounding Volume Hierarchies. In *Proc. Conference on High Performance Graphics*, pages 7–13, New York, NY, USA, 2009. ACM. 16

[SFES07] Lars Schjøth, Jeppe Revall Frisvad, Kenny Erleben, and Jon Sporring. Photon differentials. In *Proc. Computer graphics and interactive techniques in Australia and Southeast Asia*, GRAPHITE '07, pages 179–186, New York, NY, USA, 2007. ACM. 66

[Shi90] P. Shirley. A ray tracing method for illumination calculation in diffuse-specular scenes. In *Proc. Graphics interface*, pages 205–212, Toronto, Ont., Canada, Canada, 1990. 90

[SJ09] Ben Spencer and Mark W. Jones. Into the Blue: Better Caustics through Photon Relaxation. *Eurographics 2009, Computer Graphics Forum*, 28(2):319–328, March 2009. 67

[SJ13] Ben Spencer and Mark W. Jones. Progressive photon relaxation. *ACM Trans. Graph.*, 32(1):7:1–7:11, 2013. 67

[SKBF12] K. Schwenk, A. Kuijper, J. Behr, and D.W. Fellner. Practical noise reduction for progressive stochastic ray tracing with perceptual control. *Computer Graphics and Applications, IEEE*, 32(6):46 –55, nov.-dec. 2012. 80, 122

[SP89] F. Sillion and C. Puech. A general two-pass method integrating specular and diffuse reflection. *SIGGRAPH Comput. Graph.*, 23(3):335–344, July 1989. 90

[SSK07] Maxim Shevtsov, Alexei Soupikov, and Alexander Kapustin. Highly Parallel Fast KD-tree Construction for Interactive Ray Tracing of Dynamic Scenes. *Comput. Graph. Forum*, 26:395–404(10), September 2007. 16

[SSO08] Lars Schjøth, Jon Sporring, and Ole Fogh Olsen. Diffusion based photon mapping. *Comput. Graph. Forum*, 27(8):2114–2127, 2008. 66

[SSSK04] László Szécsi, Mateu Sbert, and László Szirmay-Kalos. Combined correlated and importance sampling in direct light source computation and environment mapping. *Comput. Graph. Forum*, 23(3):585–594, 2004. 40

[Sta99] Jos Stam. Diffraction shaders. In *SIGGRAPH*, pages 101–110, 1999. 8

[Suy02] Frank Suykens. *On robust Monte Carlo algorithms for multi-pass global illumination.* PhD thesis, Department of Computer Science, K.U.Leuven, Leuven, Belgium, September 2002. Willems, Yves (supervisor). 90

[SW99] Frank Suykens and Yves Willems. Combining Bidirectional Path Tracing and Multipass Rendering. In *Proc. International Conference in Central Europe on Computer Graphics, Visualization and Interactive Digital Media*, pages 265–272, 1999. 90

[SW00] Frank Suykens and Yves D. Willems. Adaptive filtering for progressive monte carlo image rendering. In *WSCG*, pages 220–227, Feb 2000. 78

[TCE05] Justin F. Talbot, David Cline, and Parris K. Egbert. Importance resampling for global illumination. In Kavita Bala and Philip Dutré, editors, *Rendering Techniques 2005 Eurographics Symposium on Rendering*, pages 139–146, Aire-la-Ville, Switzerland, 2005. Eurographics Association. 35

[TM98] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *Proc. International Conference on Computer Vision*, ICCV '98, pages 839–, Washington, DC, USA, 1998. IEEE Computer Society. 79

[TM06] R. Tobler and S. Maierhofer. Improved illumination estimation for photon maps in architectural scenes. *Proc. WSCG*, pages pp. 257–262, 2006. 66

[TM10] Daniel Weiskopf Thomas Muller, Sebastian Grottel. Special relativistic visualization by local ray tracing. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1243–1250, November 2010. 9

[TS67] K. E. Torrance and E. M. Sparrow. Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America*, 57(9):1105–1112, Sep 1967. 20

[Tsa09] John A. Tsakok. Faster Incoherent Rays: Multi-BVH Ray Stream Tracing. In *Proc. Conference on High Performance Graphics*, pages 151–158, New York, NY, USA, 2009. ACM. 17

[Tuc07] V.V. Tuchin. *Tissue Optics: Light Scattering Methods and Instruments for Medical Diagnosis*. Press Monographs. SPIE Press, 2007. 7

[TV08] Y.M. Timofeev and A.V. Vasilev. *Theoretical Fundamentals of Atmospheric Optics*. Cambridge International Science Publishing, Limited, 2008. 7

[Vea97] Eric Veach. *Robust Monte Carlo Methods for Light Transport Simulation*. PhD thesis, Stanford University, December 1997. 55, 57, 59

[VG94] Eric Veach and Leonidas Guibas. Bidirectional estimators for light transport. In *Proc. Eurographics Workshop on Rendering*, pages 147–162. Eurographics, June 1994. 54, 56

[VG95] Eric Veach and Leonidas J. Guibas. Optimally combining sampling techniques for monte carlo rendering. In *Proceedings of SIGGRAPH '95, Annual Conference Series*, SIGGRAPH '95, pages 419–428, New York, NY, USA, 1995. ACM.

[VG97] Eric Veach and Leonidas J. Guibas. Metropolis Light Transport. In *Proceedings of SIGGRAPH '97, Annual Conference Series*, pages 65–76, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co. 57, 74

[War91] Greg Ward. Real pixels. In *Graphics Gems II*, pages 80–83. Academic Press, 1991. 61

[War92] Gregory J. Ward. Measuring and modeling anisotropic reflection. *SIGGRAPH Comput. Graph.*, 26(2):265–272, July 1992. 20

[WBB08] Ingo Wald, Carsten Benthin, and Solomon Boulos. Getting Rid of Packets - Efficient SIMD Single-Ray Traversal Using Multi-Branching BVHs. In *IEEE Symposium on Interactive Ray Tracing*, pages 49–57, 2008. 17

[WBS07] Ingo Wald, Solomon Boulos, and Peter Shirley. Ray Tracing Deformable Scenes using Dynamic Bounding Volume Hierarchies. *ACM Trans. Graph.*, 26(1):6, 2007. 16

[WCG87] John R. Wallace, Michael F. Cohen, and Donald P. Greenberg. A two-pass solution to the rendering equation: A synthesis of ray tracing and radiosity methods. In *Proc. Computer graphics and interactive techniques*, SIGGRAPH '87, pages 311–320, New York, NY, USA, 1987. ACM. 90

[WFA+05] Bruce Walter, Sebastian Fernandez, Adam Arbree, Kavita Bala, Michael Donikian, and Donald P. Greenberg. Lightcuts: a scalable approach to illumination. *ACM Trans. Graph.*, 24(3):1098–1107, July 2005. 73

[WGBK07] Ingo Wald, Christian P Gribble, Solomon Boulos, and Andrew Kensler. SIMD Ray Stream Tracing - SIMD Ray Traversal with Generalized Ray Packets and On-the-fly Re-Ordering. Technical Report UUSCI-2007-012, SCI Institute, University of Utah, 2007. 17

[WGS04] Ingo Wald, Johannes Günther, and Philipp Slusallek. Balancing considered harmful – faster photon mapping using the voxel volume heuristic. *Comput. Graph. Forum*, 22(3):595–603, 2004. 62, 102

[WH06] Ingo Wald and Vlastimil Havran. On Building Fast kD-trees for Ray Tracing, and on doing that in O(N log N). In *Proc. IEEE Symposium on Interactive Ray Tracing*, pages 61–69, 2006. 16

[WH08] Gregory J. Ward and Paul S. Heckbert. Irradiance gradients. In *ACM SIGGRAPH courses*, SIGGRAPH '08, pages 72:1–72:17, New York, NY, USA, 2008. ACM. 77

[Whi80] Turner Whitted. An Improved Illumination Model for Shaded Display. *Commun. ACM*, 23(6):343–349, 1980. 22

[WIP08] Ingo Wald, Thiago Ize, and Steven G. Parker. Fast, Parallel, and Asynchronous Construction of BVHs for Ray Tracing Animated Scenes. *Computers & Graphics*, 32(1):3 – 13, 2008. 16

[WKB12] Bruce Walter, Pramook Khungurn, and Kavita Bala. Bidirectional lightcuts. *ACM Trans. Graph.*, 31(4):59:1–59:11, July 2012. 73

[WLL+09] Tim Weyrich, Jason Lawrence, Hendrik P. A. Lensch, Szymon Rusinkiewicz, and Todd Zickler. Principles of appearance acquisition and representation. *Found. Trends. Comput. Graph. Vis.*, 4(2):75–191, February 2009. 20

[WRC88] Gregory J. Ward, Francis M. Rubinstein, and Robert D. Clear. A ray tracing solution for diffuse interreflection. *SIGGRAPH Comput. Graph.*, 22(4):85–92, June 1988. 77

[WS01] Ingo Wald and Philipp Slusallek. State-of-the-Art in Interactive Ray Tracing. In *Eurographics State of the Art Reports*, pages 21–42, 2001. 3

[WW11] Alexander Wilkie and Andrea Weidlich. How to write a polarisation ray tracer. In *SIGGRAPH Asia 2011 Courses*, SA '11, pages 8:1–8:36, New York, NY, USA, 2011. ACM. 9

[WW12] Alexander Wilkie and Andrea Weidlich. Polarised light in computer graphics. In *SIGGRAPH Asia 2012 Courses*, SA '12, pages 8:1–8:87, New York, NY, USA, 2012. ACM. 9

[XP05] Ruifeng Xu and Sumanta N. Pattanaik. A novel monte carlo noise reduction operator. *IEEE Comput. Graph. Appl.*, 25(2):31–35, March 2005. 80

[YIC+10] Yonghao Yue, Kei Iwasaki, Bing-Yu Chen, Yoshinori Dobashi, and Tomoyuki Nishita. Unbiased, adaptive stochastic sampling for rendering inhomogeneous participating media. *ACM Trans. Graph.*, 29(6):177:1–177:8, December 2010. 7

[YSLH11] Lei Yang, Pedro V. Sander, Jason Lawrence, and Hugues Hoppe. Antialiasing recovery. *ACM Transactions on Graphics*, 30(3), 2011. 157

[ZERB05] Todd Zickler, Sebastian Enrique, Ravi Ramamoorthi, and Peter N. Belhumeur. Reflectance sharing: Image-based rendering from a sparse set of images. In *Rendering Techniques*, pages 253–264, 2005. 20

[ZHWG08] Kun Zhou, Qiming Hou, Rui Wang, and Baining Guo. Real-time Kd-tree Construction on Graphics Hardware. In *SIGGRAPH Asia '08*, pages 1–11, New York, NY, USA, 2008. ACM. 16

[ZL09] Zhengyun Zhang and M. Levoy. Wigner distributions and how they relate to the light field. In *Computational Photography (ICCP), 2009 IEEE International Conference on*, pages 1 –10, april 2009. 8