Chapter 12

Conclusion

We conclude by summarizing the main results of this dissertation, in somewhat more detail than they were described in Chapter 1.

12.1 Bidirectional light transport theory

In the first part of this thesis, we have investigated the theoretical basis of bidirectional light transport algorithms. We proposed two different linear operator formulations of light transport, based on different sets of assumptions. First we considered the general case, where no assumptions are made about the physical validity of the scattering models used. In this case, we cannot rely on the properties of light transport in the real world: for example, energy might not be conserved. Nevertheless, there is still a well-defined mathematical problem to be solved (with mild restrictions discussed in Chapter 4), and we describe the manipulations that are necessary to ensure that algorithms based on radiance transport, importance transport, light particles, and importance particles all converge to the same mathematically correct solution. We have given a detailed analysis of the framework, including the norms, inverses, and adjoints of the various transport operators. We have also given explicit rules for handling all the various combinations of incident and exitant quantities.

We have shown that the above model is useful whenever the scene contains materials whose bidirectional scattering distribution function (BSDF) is not symmetric. There are two distinct situations where this can arise. First, some scattering models in computer graphics are not physically valid. It is important to be able to handle non-physical materials correctly, since they are sometimes very convenient. As a particular example we consider the use of shading normals, which are commonly applied to make polygonal surfaces look smooth or to add detail to coarse geometric models. We show that shading normals modify BSDF's and make them non-symmetric. However, using the linear operator formulation above we show that it is still possible to handle shading normals correctly and consistently in bidirectional algorithms, by using the correct adjoint BSDF (which we derive).

We also show that non-symmetric BSDF's can arise even for materials that are physically valid. This occurs whenever light is transmitted between two substances with different indices of refraction. Again we show how to handle this situation correctly within the general framework above, by deriving the adjoint BSDF for refraction. The use of this adjoint BSDF is necessary to ensure that bidirectional algorithms will converge to correct results.

However, when all materials in the scene model are physically valid, we have shown that there is a much better way to formulate the light transport problem. This formulation is based on a new reciprocity principle that holds for materials that transmit as well as reflect light. In particular, for physically valid materials we have shown that it is not the BSDF $f_{\mathrm{s}}(\omega_{\mathrm{i}} \to \omega_{\mathrm{o}})$ that is symmetric, but instead the quantity $f_{\mathrm{s}}(\omega_{\mathrm{i}} \to \omega_{\mathrm{o}})/\eta_{\mathrm{o}}^2$ (where η_{o} is the refractive index of the medium containing ω_{o}). We establish this principle using the laws of thermodynamics, in particular Kirchhoff's laws and the principle of detailed balance. These laws hold for systems in thermodynamic equilibrium, but the resulting reciprocity principle is valid generally. We have investigated the historical origins of such principles, including Helmholtz and Rayleigh reciprocity, and clarified the important point that Helmholtz himself did not make any statement that would imply the symmetry of BRDF's. We have also discussed the subtle issues that arise in justifying such principles: the roles of thermodynamic equilibrium, time reversal invariance, and detailed balance. Finally, we have considered the conditions under which reciprocity does not hold, i.e. in the presence of absorbing media or external magnetic fields.

Taking advantage of this reciprocity principle, we have proposed a new light transport model where the transport operators are symmetric (self-adjoint) for any physically valid scene model. This symmetry simplifies both the theory of light transport algorithms (by eliminating the need for adjoint operators), and also their implementation (since the same transport rules apply to light and importance, or to path tracing and particle tracing). Furthermore, the transport quantities in the new model are optical invariants, which creates interesting connections to classical geometric optics. The modifications relative to the previous formulation are straightforward, and simply involve scaling the various transport quantities by the square of the refractive index of the surrounding medium. We have also provided a detailed analysis of the norms, inverses, and adjoints of the new operators.

We have proposed a third theoretical model for light transport, where each measurement is expressed as an integral over a space of paths (rather than as the solution to an integral equation or linear operator equation). The main advantage of this approach is its simple abstract form: by reducing light transport to a set of integrals, it allows general-purpose integration and sampling techniques to be applied (such as multiple importance sampling, or the Metropolis method). It is also useful from a conceptual point of view, since this formulation makes it clear that paths can be sampled in virtually any way desired, not just by recursively sampling a transport equation. We have described a variety of natural measures on paths with well-defined physical meanings, and we have developed an extended regular expression notation for paths that describes the properties of sources and sensors as well as the scattering properties at intermediate vertices. We have used this model to analyze the capabilities of unbiased Monte Carlo sampling algorithms, and we have shown that there are certain kinds of paths that cannot be generated by standard sampling techniques. This implies that certain lighting effects will be missing from the images generated using these techniques. We have analyzed the conditions under which this occurs, and we have proposed methods for making these path sampling algorithms complete.

12.2 General-purpose Monte Carlo techniques

The second area of this dissertation concerns new general-purpose techniques for Monte Carlo integration. Our main contribution in this area is *multiple importance sampling*, a method for combining several different sampling techniques for the same integral in order to obtain low-variance estimators for a broad class of integrands. We started by proposing a general model for combining samples from different techniques, called the *multi-sample* *model*. Using this model, we showed that any unbiased combination strategy could be represented as a set of weighting functions. This gave us a large space of possible combination strategies to explore, and a uniform way to represent them. We then proposed a specific combination strategy called the *balance heuristic*, and we proved that the variance obtained using this strategy is optimal to within a small additive term. We then proposed several other combination strategies, which are basically refinements of the balance heuristic: their variance is also provably close to optimal, but they give better results in certain important special cases.

We tested these methods on a variety of integration problems in computer graphics, and we found that multiple importance sampling can reduce variance substantially at little extra cost. The method is simple and practical to implement, and can make Monte Carlo calculations significantly more robust.

We have also proposed a new technique called *efficiency-optimized Russian roulette*. We started by showing that the variance of Russian roulette can be analyzed as a function of its threshold parameter (whose value is usually chosen in an *ad hoc* manner). We then described a technique for choosing the value of this parameter in order to maximize the efficiency of the resulting estimator. The main application of this technique in graphics is to reduce the number of visibility tests in rendering problems.

12.3 Robust light transport algorithms

We have shown how these theories and techniques can be applied to the construction of robust Monte Carlo light transport algorithms. The first algorithm we described was *bidirectional path tracing*, which is based on the path integral framework: it generates paths using a family of different importance sampling techniques, and then combines them using multiple importance sampling. Specifically, each path is constructed by concatenating two subpaths, one generated starting from a light source and another generated starting from the camera. We have shown that each such technique can efficiently sample a different set of paths, and that these paths are responsible for different lighting effects in the final image. By combining samples from all the techniques, we can efficiently render scenes under a wide variety of illumination conditions.

In addition to describing the mathematical basis of the method, we also discussed the implementation issues in detail. This includes how to sample and filter of the image, how to generate the paths and evaluate their contributions efficiently, and how to implement the important special cases where the light or eye subpath contains at most one vertex. We have also described the extensions required to handle ideal specular surfaces, and we have shown how efficiency-optimized Russian roulette can be used to reduce the number of visibility tests.

Bidirectional path tracing is unbiased, straightforward to implement, and supports the same range of geometry and materials as standard path tracing. It is an effective rendering algorithm for many kinds of indoor scenes, and is particularly useful for scene models with concentrated indirect lighting. On the other hand, the main weakness of the algorithm is that the light and eye subpaths are generated independently. This makes it unsuitable for outdoor environments, or scenes with many light sources, or scenes with constricted geometry between the light sources and the viewer.

Finally, we have introduced a new algorithm called *Metropolis light transport*. This method is also based on the path integral framework, but it samples paths in a different way. Specifically, it uses the Metropolis sampling algorithm, which generates a sequence of paths by following a random walk through path space. Each path is generated from the previous one by proposing a random mutation. This mutation is then either accepted or rejected with a carefully chosen probability, in order that the probability density of sampling each path is proportional to the contribution it makes to the desired final image. The resulting algorithm is unbiased, handles general geometric and scattering models, and can be far more efficient than previous algorithms on scenes with complex illumination. Furthermore, it is competitive with previous unbiased algorithms even for scenes whose lighting is relatively simple.

To derive this method we first proposed a slight modification to the path integral framework that allows paths to be sampled across the entire image (rather than within each pixel separately). We showed that the Metropolis algorithm can then be used to determine the relative pixel intensities across the image, while the overall image brightness needs to be determined during a separate initialization phase. We addressed the issue of start-up bias during initialization (a common problem with Metropolis applications), and showed that in the case of light transport this bias can be eliminated completely. For the Metropolis phase of the algorithm, we proposed a set of criteria for designing path mutations in order to minimize the error in the final image. We have also described three different mutation strategies we have implemented that partially satisfy these goals, namely *bidirectional mutations*, *lens subpath mutations*, and *perturbations*. Finally, we described several refinements to the basic algorithm that improve its performance in practice.

The main advantage of Metropolis light transport is its ability to handle complex illumination efficiently, by exploring the space of paths that actually contribute to the image. Unlike bidirectional path tracing, it can also handle problems where only a small fraction of the emitted light in the scene reaches the viewer (e.g. due to difficult visibility). Furthermore, since it is a Monte Carlo algorithm it can support complex geometry and materials efficiently. We feel that the ability to handle complex geometry, materials, and illumination is an important goal, since light transport algorithms need to produce reliable, consistent results over the widest possible range of real environments if they are ever going to be widely used.