

Optimization of Illumination through Windows for MCRT

Ildar Valiev¹, Dmitry Zhdanov², Sergey Ershov¹, Alexey Voloboy¹, Vadim Sokolov¹, Vladimir Galaktionov¹

piv@gin.keldysh.ru|ddzhdanov@mail.ru|sergey_65@mail.ru|voloboy@gin.keldysh.ru|sokolov1969@gmail.com|vlgal@gin.keldysh.ru

¹The Keldysh Institute of Applied Mathematics Russian Academy of Sciences, Moscow, Russia

²ITMO University, St. Petersburg, Russia

The paper presents an improvement of Monte-Carlo ray tracing which changes ray emission from a light source to accelerate convergence i.e. reduce the noise remained after the given run time. It is mainly intended for interior scenes illuminated from outside (e.g. skylight) through windows or other holes. The rays from light sources are generated so that they are directed to these windows. In other words, the number of rays is increased for directions that contribute to the camera image. It is shown that the proposed method allows calculating image with desirable quality several times faster.

Keywords: ray tracing, lighting simulation, Monte-Carlo, optimal PDF.

1. Introduction

The basic Monte-Carlo ray tracing (MCRT) generates or transforms rays at random with probability distribution determined locally i.e. in case of emission by the light source properties and in case of surface or particle scattering by Bidirectional scattering function (BSDF) in the hit point.

This approach is not ultimately optimal, e.g. BSDF may send rays away from the virtual observer or camera. Moreover, in case of a parallel light source there is an ambiguity with the ray origin: formally for a parallel light the ray origin must be chosen uniformly in an infinite domain which is technically impossible. Therefore in practice the ray origin is chosen over the projection of scene onto the plane orthogonal to the ray direction (see Fig. 1).

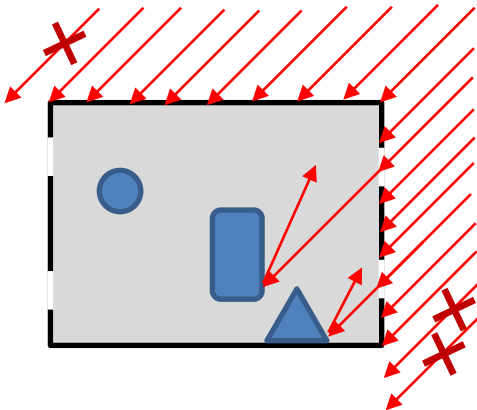


Fig. 1. Top view of a scene illuminated by a parallel light source. Rays from the light are emitted to the scene only.

Rays that miss the window do not contribute to the interior view.

Therefore the ray emission is determined not only by the light source properties but also by the geometry of scene. The natural extension of that approach is when we render an interior of a room illuminated from outside through a window (or an open door). In this case the rays sent to the opaque walls are “lost” while illumination of the interior is created only by the rays that run through the window. The ray origin must then be chosen in a projection of that window onto the plane orthogonal to the illumination direction.

This is the basic idea behind optimization of light emission.

Below we shall describe how it must be adapted to the more complex cases, when

- Light source is not parallel but it is skylight with given goniogram;
- There are several windows that can overlap.

The examples shown in Fig. 1 and 2 can be implemented by letting the user to point the windows and then mask the distribution of ray origin by 1 inside or 0 outside their projection. This approach, however, may introduce error. For example, imagine a house with “French windows” stayed on a white sand lawn. The sun rays that hit the ground near the window can reflect into it and contribute to the illumination of the interior. Meanwhile, the light emits only towards the transparent windows thus no rays are directed to the opaque ground (Fig. 2).

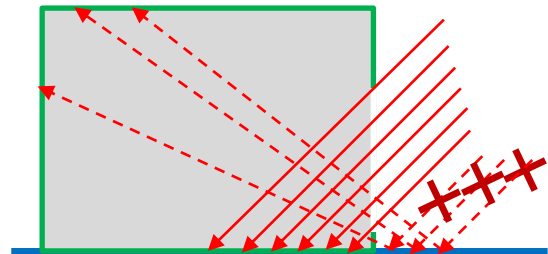


Fig. 2. Side view of a scene illuminated by a parallel light source. Rays are directed to the windows only and not to the ground. These rays (shown dashed) can reflect inside the room too but are missed.

This situation can be cured if the “emission mask” is more sophisticated and is not 0 for some opaque areas. On the contrary, the mask is not 1 for some transparent areas even they do not open into the interior.

The resulting Probability density function or PDF of the emitted ray origin (and/or direction) cannot be simply postulated but must be calculated from the results of simulation. The method then belongs to a wide class of algorithms of an “optimal PDF” for MCRT.

2. Relation to previous work

Usually the quantities of interest for optical simulation, e.g. the luminance of the hit point, are usually the integrals of a product, e.g. the point luminance is a convolution of BSDF with

illumination. Therefore the optimal probability distribution of the scattered ray direction in MCRT must be that product. This is usually termed as importance sampling. The problem is that at one multiplier of that product (illumination in the above example) is the result of MCRT itself and is not known in advance and accurately.

Since long attempts are made to improve efficiency of MCRT by making the PDF of scattered rays as close as possible to that product using various approximations and heuristic. Say, surface luminance under even direct illumination by an area light is an integral over that light surface. To improve efficiency of its calculation Shirley et al [1] suggested to separate the fastest varying terms of the integrand and to use them for constructing a PDF.

The famous works [2], [3] improved efficiency of the backward MCRT by scattering the camera ray according to not BSDF only but its product with the illumination of the point. In [2] photon maps were used to estimate the angular distribution of illuminance. In [3] this was made with a 5D tree covering the scene which accumulated illumination brought by various reflected backward MCRT rays (so there was no a separate forward MCRT phase).

There are two main problems with the method. Firstly the optimal distribution must now be stored if not in each hit point but in each voxel. This requires a lot of memory. Various approximations are used to reduce it, e.g. the space resolution is quite coarse. And this results in the second problem: we have only a rather rough estimation of the incoming radiation distribution. It thus deviates from the target product and the difference must be compensated by the scale of ray energy.

At last, we may face a deadlock with the methods like [3] where the illumination is taken from the same MCRT rays which were shot with the accumulated PDF. For example, if by some time no rays were shot in some angular cone the estimated illumination from it is 0. And when generating a next ray the sampling procedure does not send it to that “black” region, so illumination from it remains 0 and so on. The simplest remedy is to use a mixed strategy when some fraction of rays are scattered by BSDF without usage of the accumulated PDF. The difficulty though is how to find the ratio. If it is too small the narrow peaks of illumination do not contribute to PDF and are only taken from the small fraction of rays scattered by BSDF which creates strong noise. If the ratio is large this kills the very benefit (acceleration) of the method.

This basic idea was then exploited by many researchers [4]–[8]. While [5] uses hemispherical PDF of the form rather similar to Jensen’s [2], in [6] the authors use approximation by a sum of Gaussians. Its advantage is that a product of two Gaussian bells is also a Gaussian and can be sampled efficiently. In [8] wavelets are used instead. Some authors utilize spherical harmonics whose product also admits analytic treatment. Also other approaches are used up to neural networks and machine learning which are another way to optimize the ray direction PDF on the base of accumulated statistic [7].

All the above methods alter PDF in one point to reduce variance. A radical solution is the Metropolis Light transport where, unlike MCRT which traces ray segment-by segment, we choose at random the whole path from light source to camera [9]. But while the above problem of MCRT is eliminated other problems emerge whose analysis is out of our scope.

Less radical solution is still using MCRT but applying different approaches (both for generation of scattering direction and for collecting illuminance) and generating several ray paths. The pixel luminance is then taken as a weighted sum of luminance each of them would brought. This is usually termed as multiple importance strategy. This method does not optimize PDF in hit points separately but instead it mixes contributions of several different

whole ray paths. The main ideas were suggested in [10] where besides all it was proved that the relatively simple “balance heuristic” estimator is close to optimal, i.e. other estimators of the class considered cannot decrease the variance significantly.

Recently Sbert et al [11] claimed this is not always so. They analyzed the conditions in the Veach’s theorem and found that sometimes an estimator can be constructed significantly better than the “balance heuristic” one.

3. Light entrance through the windows

3.1 Simple transparent window

For a parallel light source the ray origin is distributed uniformly. On the other hand rays that miss the window are useless because they do not create interior illumination. Therefore the ray origin is chosen uniformly within projection of the window onto the plane orthogonal to the light direction.

In case of triangulated scene geometry the triangles that constitute the window glass are used. First we choose one of them with probability proportional to the energy flow through that triangle and then choose the ray origin uniformly within the triangle.

Although it is possible that the window is open or does not have glass inside. In this case there are no triangles inside the window to direct rays to. The user then can select triangles that form the window’s frame. After that a convex hull is constructed from them and the rays are directed to its triangles. In case light source is not parallel the calculation is a bit more sophisticated. Say, sky light, like a parallel light, is an infinitely distant, but it has a smooth goniogram.

Here we first choose direction of emission \mathbf{v} and after that we choose triangle t with probability

$$P_t = \frac{S_t |(\mathbf{v} \cdot \mathbf{n}_t)|}{\sum_t S_t |(\mathbf{v} \cdot \mathbf{n}_t)|} \quad (1)$$

where S_t is area of the t -th triangle and \mathbf{n}_t is its normal. Finally the ray origin \mathbf{x} is chosen uniformly in that triangle and translated outside the scene domain. Direction \mathbf{v} is chosen with probability density

$$p(\mathbf{v}) = \frac{\sum_t S_t |(\mathbf{v} \cdot \mathbf{n}_t)| g(\mathbf{v})}{\sum_t S_t \int |(\mathbf{v} \cdot \mathbf{n}_t)| g(\mathbf{v}) d^2 \mathbf{v}}$$

where g is intensity of the emission goniogram.

Let us assume that sky goniogram is tabulated on a rectangular grid (ϑ, φ) , and bilinearly interpolates inside cells.

We first choose goniogram cell $[\vartheta_i, \vartheta_{i+1}] \times [\varphi_j, \varphi_{j+1}]$ with probability

$$P_{i,j} = \frac{\sum_t S_t \int_{cell_{i,j}} |(\mathbf{v} \cdot \mathbf{n}_t)| g(\mathbf{v}) d^2 \mathbf{v}}{\sum_t S_t \int |(\mathbf{v} \cdot \mathbf{n}_t)| g(\mathbf{v}) d^2 \mathbf{v}} \quad (2)$$

and then choose within this cell a direction with probability density

$$p_{i,j}(\mathbf{v}) = \frac{\sum_t S_t |(\mathbf{v} \cdot \mathbf{n}_t)| g(\mathbf{v})}{\sum_t S_t \int_{cell_{i,j}} |(\mathbf{v} \cdot \mathbf{n}_t)| g(\mathbf{v}) d^2 \mathbf{v}} \quad (3)$$

After that we choose triangle t with probability (1).

If the cell is small so that the emission directions for all its 4 vertices are to the same side of triangle normal then the integral over cell is

$$\int_{cell_{i,j}} |(\mathbf{v} \cdot \mathbf{n}_t)| g(\mathbf{v}) d^2 \mathbf{v} = |(\mathbf{F}_{i,j} \cdot \mathbf{n}_t)|$$

where

$$F_{i,j} \equiv \int_{\text{cell}_{i,j}} \mathbf{v} g(\mathbf{v}) d^2 \mathbf{v}$$

depends only on the goniogram cell (but not triangle). Therefore for a goniogram cell it suffices to calculate three values $F_{i,j}^{(x)}, F_{i,j}^{(y)}, F_{i,j}^{(z)}$ and then for most of triangles the flow through them is calculated as a dot product of this vector with triangle normal. For those few triangles which are illuminated from different sides by the goniogram cell the energy flow must be calculated directly.

Therefore in the above method the probability density of the ray origin and direction (\mathbf{x}, \mathbf{v}) is

$$p(\mathbf{x}, \mathbf{v}) = \frac{\sum_t \chi_t(\mathbf{x}; \mathbf{v}) g(\mathbf{v})}{\int \sum_t \chi_t(\mathbf{x}; \mathbf{v}) g(\mathbf{v}) d^2 \mathbf{v} d^2 \mathbf{x}} \quad (4)$$

$$= \frac{\sum_t \chi_t(\mathbf{x}; \mathbf{v}) g(\mathbf{v})}{\sum_t S_t \int |(\mathbf{v} \cdot \mathbf{n}_t)| g(\mathbf{v}) d^2 \mathbf{v}}$$

where $\chi_t(\mathbf{x}; \mathbf{v}) = 1$ if the ray (\mathbf{x}, \mathbf{v}) intersects triangle t and 0 otherwise.

3.2 Multiple windows and overlap

The proposed above method produces correct energy flow through a surface composed by triangles (or other facets) whose projections do not overlap. If projections of N triangles overlap the numerator in (4) increases for this area and this causes a distortion, see Fig. 3.

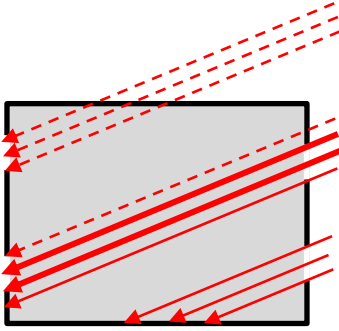


Fig. 3. Top view of a scene with multiple windows. The rays shown with thick arrows intersect two windows. Dashed arrows show rays directed to windows that were shadowed by an opaque surface before entering the room.

The obvious correction is to change (4) to

$$p(\mathbf{x}, \mathbf{v}) = \frac{\sum_t N^{-1}(\mathbf{x}; \mathbf{v}) \chi_t(\mathbf{x}; \mathbf{v}) g(\mathbf{v})}{\int \sum_t N^{-1}(\mathbf{x}; \mathbf{v}) \chi_t(\mathbf{x}; \mathbf{v}) g(\mathbf{v}) d^2 \mathbf{v} d^2 \mathbf{x}} \quad (5)$$

where $N(\mathbf{x}; \mathbf{v}) = \sum_t \chi_t(\mathbf{x}; \mathbf{v})$ is the number of triangles the ray $(\mathbf{x}; \mathbf{v})$ intersects.

This correcting factor is easily handled with the “rejection method”: we first choose ray direction and origin according to the probability density (4), i.e. actually with (1), (2) and (3). Then once cycles over all triangles of all windows and check if the ray intersects them. This is done for isolated triangle, i.e. we do not care if there are other triangles “in front of it”. Then with probability $1 - N^{-1}(\mathbf{x}; \mathbf{v})$ the ray is rejected, i.e. the direction and origin are chosen anew, and this continues until the ray is accepted. The resulting density will then be (4) then it provides correct illumination.

4. Adaptive optimal PDF

As seen from (5) the above approach masks the own spatially uniform light source probability density $g(\mathbf{v})$ with mask which is 1 if the ray intersect any window triangle and 0 otherwise. This simplest alteration is not the optimal. Let us consider a more flexible approach when the origin and direction of the emitted ray are chosen with arbitrary probability density $\rho(\mathbf{x}; \mathbf{v})$.

After the light source ray $(\mathbf{x}; \mathbf{v})$ is chosen its further fate is still stochastic at each hit point. So its contribution to the luminance of pixel p is random function $L_p(\mathbf{x}; \mathbf{v})$. Naturally if one changes the distribution of rays it must be compensated by the change of the ray contribution to keep the expectation. Therefore

$$L_p(\mathbf{x}; \mathbf{v}) = \mathcal{L}_p(\mathbf{x}; \mathbf{v}) \frac{p(\mathbf{x}; \mathbf{v})}{\rho(\mathbf{x}; \mathbf{v})}$$

where $\mathcal{L}_p(\mathbf{x}; \mathbf{v})$ is contribution for the correct density (5).

Assuming the camera ray path is deterministic, the noise in pixel p is entirely due to the forward MCRT part and can be calculated as

$$\int \langle \mathcal{L}_p^2 \rangle(\mathbf{x}; \mathbf{v}) \rho(\mathbf{x}; \mathbf{v}) d^2 \mathbf{x} d^2 \mathbf{v} - \left(\int \langle \mathcal{L}_p \rangle(\mathbf{x}; \mathbf{v}) \rho(\mathbf{x}; \mathbf{v}) d^2 \mathbf{x} d^2 \mathbf{v} \right)^2$$

where $\langle \cdot \rangle$ means averaging over all random fates of the ray emitted from light source with origin \mathbf{x} and direction \mathbf{v} . The total error is just the sum over all pixels.

Varying the probability density by $\delta\rho$ changes this total error by

$$- \int \sum_p \langle \mathcal{L}_p^2 \rangle(\mathbf{x}; \mathbf{v}) \frac{p^2(\mathbf{x}; \mathbf{v})}{\rho^2(\mathbf{x}; \mathbf{v})} \delta\rho(\mathbf{x}; \mathbf{v}) d^2 \mathbf{x} d^2 \mathbf{v} \quad (6)$$

An optimal $\rho(\mathbf{x}; \mathbf{v})$ makes the noise minimal and in this extremum the change must vanish for any admissible $\delta\rho$. The variation of normalized density is by definition an arbitrary function for which $\int \delta\rho(\mathbf{x}; \mathbf{v}) d^2 \mathbf{x} d^2 \mathbf{v} = 0$. Comparing with (6) we thus conclude that it must be (the common derivation in Monte-Carlo methods)

$$\sum_p \langle \mathcal{L}_p^2 \rangle(\mathbf{x}; \mathbf{v}) \frac{p^2(\mathbf{x}; \mathbf{v})}{\rho^2(\mathbf{x}; \mathbf{v})} = \text{const}$$

or

$$\rho(\mathbf{x}; \mathbf{v}) = C p(\mathbf{x}; \mathbf{v}) \sqrt{\sum_p \langle \mathcal{L}_p^2 \rangle(\mathbf{x}; \mathbf{v})} \quad (7)$$

where C is the normalization constant.

The $\langle \mathcal{L}_p^2 \rangle(\mathbf{x}; \mathbf{v})$ is by definition independent on how we emit the ray $(\mathbf{x}; \mathbf{v})$. Before ray tracing we create a 4D mesh in that $(\mathbf{x}; \mathbf{v})$. Then when a ray is emitted we find which cell the $(\mathbf{x}; \mathbf{v})$ belongs to and add $\mathcal{L}_p^2(\mathbf{x}; \mathbf{v})$ to the value accumulated in that cell.

As ray tracing continues we accumulate better and better estimation of $\langle \mathcal{L}_p^2 \rangle(\mathbf{x}; \mathbf{v})$ in each cell and begin using it to optimize emission density. The ray origin and direction are now chosen with density (7). As ray tracing continues and the accumulated estimate of $\langle \mathcal{L}_p^2 \rangle(\mathbf{x}; \mathbf{v})$ changes, the density of ray emission changes too. In other words it is now time-dependent. This however does not create a feedback loop because the accumulated value does not depend on how we choose the ray start. Gradually, as the accumulated estimate of $\langle \mathcal{L}_p^2 \rangle(\mathbf{x}; \mathbf{v})$ improves, the density of ray start converges to its limiting optimal form.

5. Results

An example scene is oversimplified model of two rooms separated by a wall with a door. Skylight illuminates the right room through the window and light penetrates into the left room through the gap beneath the door only (Fig. 4). Camera was in the left room and looks towards the door.

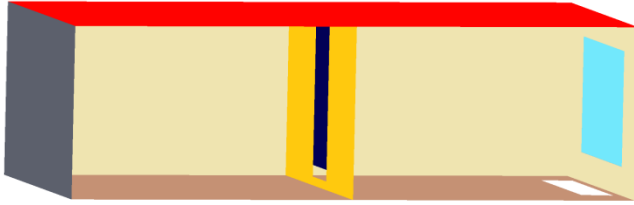


Fig. 4. Layout of the “Two rooms” scene.

Figures 5–7 present the camera images for “Two rooms” scene calculated during the same time interval without light emission optimization (Fig. 5), for light emission towards the gap beneath the door, i.e. algorithm described in p. 3 (Fig. 6), and for adaptive light PDF (Fig. 7). We also calculated average illuminance and RMS over the red rectangle presented in Fig. 5-7. The calculated values are in Table 1.

	Average	RMS
Without optimization (Fig. 5)	0.49	0.62
Light emission toward the gap (Fig. 6)	0.23	0.11
Adaptive light PDF (Fig. 7)	0.47	0.27

Table 1. Average illuminance and RMS over the red rectangle in figures 5-7 (cd/m^2).

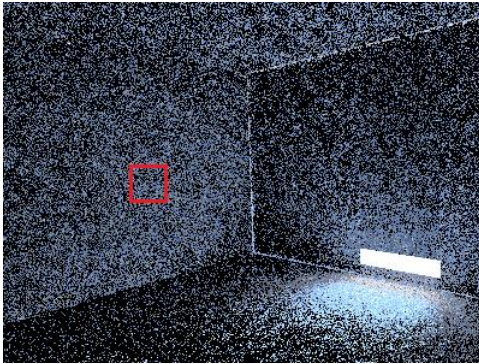


Fig. 5. “Two rooms” scene. Light emission to the whole scene domain, without optimization.

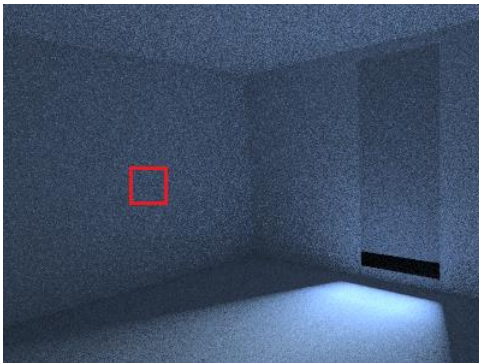


Fig. 6. “Two rooms” scene. Light emission to the gap

beneath the door.

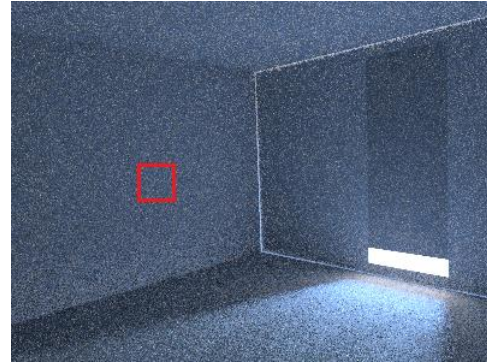


Fig. 7. “Two rooms” scene. Generation of ray direction from the light source uses an adaptive PDF

One can see that the least noise is in the second method but it underestimates luminance outside of the bright area on the floor, for example, in the wall area marked by red rectangular. Besides the gap in it looks entirely black because camera rays sent to it go to the right room which is entirely black because illumination goes to the gap only.

Another example scene is shown in Fig. 8. The house is illuminated with sky and sun light. There is a lot of windows through which it can penetrate into the interior of the room observed by the camera. Therefore the second method (explicit specification of the windows in the scene) is inconvenient here.



Fig. 8. Layout of the “House” scene.

Figures 9–10 present the images calculated during the same time interval without light emission optimization (Fig. 9) and for adaptive light PDF (Fig. 10). It is seen from these figures that level of noise is noticeably low for the adaptive light PDF method. Also visually the same quality of image was obtained about 4 times faster for the adaptive light PDF.

6. Conclusion

We suggested two methods that allow to improve ray tracing for scenes where the area of interest is illuminated from outside through windows or small holes and maybe along a complex path. Both methods reduce noise or, which is the same, allow to achieve its target level faster. The first one, which requires marking explicitly those holes in the scene geometry, gives better noise

reduction but in some cases it may underestimate illuminance as it can be seen from Table 1. The second method constructs an optimal PDF of light emission which minimizes the noise. It is slower but gives a correct illuminance. Nevertheless the second method accelerates image generation in times in comparison to light emission without optimization to the whole domain.

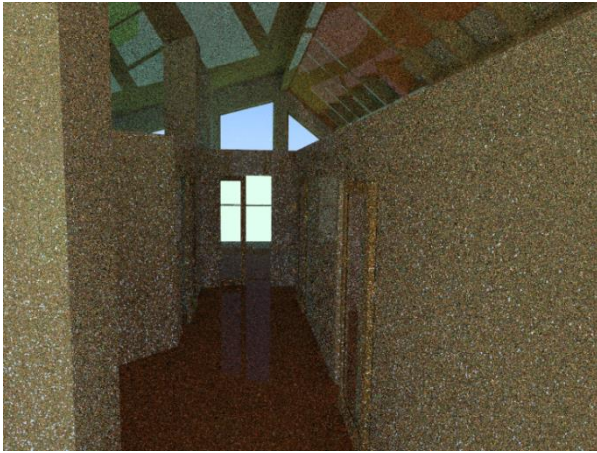


Fig. 9. Camera image for the “House” scene. Light emission to the whole scene domain.



Fig. 10. Camera image for the “House” scene. Light emission by the adaptive light PDF.

7. Acknowledgements

This work was supported by RFBR grants 18-01-00569 and 19-01-00435.

8. References

- [1] Peter Shirley, Changyaw Wang, and Kurt Zimmerman. 1996. Monte Carlo techniques for direct lighting calculations. *ACM Trans. Graph.* **15**(1), 1–36 (1996). DOI <http://dx.doi.org/10.1145/226150.226151>
- [2] Jensen, H.W. Importance driven path tracing using the photon map. In *Rendering Techniques* (proc. EGWR) (1995)
- [3] Lafortune E.P., Willems Y.D. A 5D Tree to Reduce the Variance of Monte Carlo Ray Tracing. In: Hanrahan P.M., Purgathofer W. (eds) *Rendering Techniques '95*. Eurographics. Springer, Vienna (1995). DOI https://doi.org/10.1007/978-3-7091-9430-0_2
- [4] D. Burke, A. Ghosh and W. Heidrich. Bidirectional Importance Sampling for Direct Illumination. *Eurographics Symposium on Rendering* (2005)
- [5] Heinrich Hey and Werner Purgathofer. Importance sampling with hemispherical particle footprints. In *Proceedings of the 18th Spring Conference on Computer Graphics (SCCG '02)*. ACM, New York, NY, USA, 107-114. (2002) DOI <http://dx.doi.org/10.1145/584458.584476>
- [6] Sebastian Herholz, Oskar Elek, Jiří Vorba, Hendrik Lensch, Jaroslav Křivánek. Product Importance Sampling for Light Transport Path Guiding. *Eurographics Symposium on Rendering 2016*, **35**(4) (2016)
- [7] Vorba J., Karlík O., Šik M., Ritschel T., Křivánek J.: On-line learning of parametric mixture models for light transport simulation. *ACM Trans. Graph.* (Proc. of SIGGRAPH) **33** (2014).
- [8] Petrik Clarberg, Wojciech Jarosz, Tomas Akenine-Möller, and Henrik Wann Jensen. 2005. Wavelet importance sampling: efficiently evaluating products of complex functions. *ACM Trans. Graph.* **24**, 3 (July 2005), 1166-1175. DOI: <https://doi.org/10.1145/1073204.1073328>
- [9] Veach E., Guibas L. J. Metropolis Light Transport. // *SIGGRAPH '97 Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, 65–76 (1997)
- [10] Veach E., Guibas L. J. Optimally combining sampling techniques for Monte Carlo rendering. *ACM SIGGRAPH 1995*, pp. 419–428
- [11] Mateu Sbert, Vlastimil Havran and Laszlo Szirmay-Kalos. Multiple importance sampling revisited: breaking the bounds. *EURASIP Journal on Advances in Signal Processing* (2018) 2018:15 <https://doi.org/10.1186/s13634-018-0531-2>